To

Satyavati Attayya,

Amma and Nanna

POST GRADUATE OFFICE
Submitted on 25/4/73

CERTIFICATE

Certified that this work, "Some Contributions to the Theory of Maximum Likelihood Estimation for Dependent Random Variables" by M. Siva Prasad, has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

B.L.S. Prakasa Rao
Dr. B.L.S. Prakasa Rao
Associate Professor
Department of Mathematics
Indian Institute of Technology
Kanpur

April 25, 1973.

POST GRADUATE OFFICE
This thesis has been approved
for the award of the Degree of
Doctor of Philosophy (Ph.D.)
in accordance with the
regulations of the Indian
Institute of Technology Kanpur
Dated: 8/8/73

ACKNOWLEDGEMENTS

The author wishes to express his deep gratitude to Professor B.L.S. Prakasa Rao of the Department of Mathematics for his invaluable guidance and keen interest through out this investigation.

The author also extends his thanks to Sri K.N.Tewari for his excellent maiden-performance in typing the thesis.

Financial assistance received through the Council of Scientific and Industrial Research, New Delhi is also gratefully acknowledged.

TABLE OF CONTENTS

SYNOPSIS vii

CHAPTER I INTRODUCTION 1

CHAPTER II MAXIMUM LIKELIHOOD ESTIMATION FOR STATIONARY m-DEPENDENT RANDOM VARIABLES

2.1 Introduction 10

2.2 Preliminaries 11

2.3 Regularity Conditions 16

2.4 The Asymptotic Properties of a MLE 18

2.5 Examples 30

2.6 Conclusions 40

CHAPTER III MAXIMUM LIKELIHOOD ESTIMATION FOR STATIONARY ϕ-MIXING PROCESSES

3.1 Introduction 41

3.2 Preliminaries 42

3.3 Weak Law of Large Numbers for Double Sequences of Random Variables 60

3.4 Central Limit Theorem for Double Sequences of Random Variables 73

3.5 Regularity Conditions 83

3.6 Asymptotic Properties of AMLE 86

3.7 Examples 94

3.8 Conclusions 97

CHAPTER IV MAXIMUM LIKELIHOOD ESTIMATION FOR DEPENDENT RANDOM VARIABLES

4.1 Introduction 99

4.2 Preliminaries 100

4.3 Regularity Conditions 103

4.4 Asymptotic Properties of a MLE 106

4.5 The Bernstein-vonMises Theorem 119

4.6 Examples 130

4.7 Conclusions 137

CHAPTER V COMMENTS AND SUGGESTIONS FOR FURTHER WORK 138

REFERENCES 140

SYNOPSIS

The validity of certain statistical procedures depends on the properties of the method of maximum-likelihood, viz., its consistency and the fact that it leads to an estimator whose asymptotic variance can be derived easily from the likelihood function. These properties are well established for the case where successive observations are independent and identically distributed. This will not always be the case in practice. For example, the random variables may be related by a stochastic difference equation in which case the observations are not independent. Since the error terms involved in the difference equation need not always have the same distribution, the observations may not even be identically distributed. These instances are the motivations for the study of asymptotic properties of the maximum likelihood estimators (MLE) in some general cases. Considerable work has been done in this direction to cover the case of stationary Markov processes and the case when the observations are independent but not identically distributed. In this investigation, we consider the problem of obtaining the asymptotic properties of MLE for the case of dependent random variables. This is done in three different cases, namely, when the observations are from (i) stationary

m-dependent sequences of random variables , (ii) stationary ϕ-mixing processes , and (iii) arbitrary stochastic processes.

The present thesis is divided into five chapters. A brief review of work done upto now in the area of maximum likelihood estimation relevant to our discussion, and a summary of the results obtained in this thesis are given in Chapter I.

We obtain weak consistency and first-order efficiency of MLE for stationary m-dependent random variables in Chapter II. Weak consistency is proved using the Bernstein's method of grouping random variables which he used to obtain the central limit theorem for dependent random variables. Under the set of regularity conditions presented in this chapter, we are unable to obtain asymptotic normality. However, some comments are made about the asymptotic distribution. An example, satisfying the regularity conditions , is given.

In Chapter III, this problem is considered for stationary ϕ-mixing processes. We prove a form of weak law of large numbers and central limit theorem for double sequences of random variables. Under the set of conditions presented in this chapter, these results are used to prove weak consistency and asymptotic normality of an approximate maximum likelihood estimator (AMLE) which is similar to the usual MLE. Examples satisfying the conditions are given.

A more general problem of establishing strong consistency, asymptotic normality and first-order efficiency of MLE in the case of arbitrary stochastic process, is discussed in Chapter IV. The regularity conditions are expressed in terms of probability distribution of observations conditioned upon all past observations. As an application of these results, the Bernstein-vonMises theorem regarding the convergence of posterior density to normal is proved. We use this theorem to establish strong consistency, asymptotic normality and some other asymptotic properties of Bayes estimators. Examples are provided to illustrate the theory of this chapter.

Chapter V includes some comments on our work and suggestions for further research in this important field.

CHAPTER I

INTRODUCTION AND SUMMARY

The method of maximum likelihood estimation is in common use for many years. The method can be described as follows. For a specified statistical model, given the sample values of $X_1,\ldots,X_n$, that is, $(x_1,\ldots,x_n)$, the likelihood function is a function on the parameter space Θ or on the family $\mathcal{P}$ of probability measures which, for a given value of θ (respectively P), is proportional to the (probability) density function of $(X_1, \ldots, X_n)$ at $(x_1, \ldots, x_n)$. The factor of proportionality may depend on $(x_1,\ldots,x_n)$. Thus, if $p(x_1, \ldots, x_n; \theta)$ is the joint (probability) density function of $(X_1, \ldots, X_n)$ under θ, the likelihood function of θ, given $(x_1, \ldots, x_n) = \underset{\sim}{x}_n$, is

$$\lambda(\theta; \underset{\sim}{x}_n) = K(\underset{\sim}{x}_n)\, p(x_1, \ldots, x_n; \theta), \quad \theta \in \Theta$$

A maximum likelihood estimator (MLE) of θ is defined as a point $\hat{\theta}_n$ in Θ, if such a point exists, for which $\lambda(\theta; \underset{\sim}{x}_n)$ attains its supremum. Roughly speaking, one could say that the probability of observing a value in a close neighbourhood of the sample value is maximised when the state of nature is a maximum likelihood estimator. It is to be noted here that $\lambda(\theta; \underset{\sim}{x}_n)$ does not have to be differentiable with respect to θ.

Maximum likelihood method serves as a basis for large sample procedures of testing hypotheses and of confidence interval estimation. Moreover, maximum likelihood estimators possess the invariance property, namely, if $\hat{\theta}_n$ is an MLE of θ then $g(\hat{\theta}_n)$ is an MLE of $g(\theta)$ for any function $g(\theta)$ of θ (not necessarily one-to-one). Furthermore, under some regularity conditions, maximum likelihood estimators $\hat{\theta}_n$ of θ have some desirable properties such as consistency (that is, $\hat{\theta}_n$ converges to the true parameter value θ_o of θ either in probability or almost surely as the sample size n tends to infinity), asymptotic normality (that is the asymptotic distribution of $\hat{\theta}_n$ is normal), and asymptotic efficiency (in the sense that the covariance matrix of the asymptotic distribution of $\hat{\theta}_n$ is the inverse of the Fisher information matrix).

The concept of asymptotic efficiency in the sense given above, which was called by Wald [50] as strict sense efficiency, is based on the covariance matrix of the asymptotic distribution of the estimators. Since the estimators may not have limit distributions, Wald [50] introduced the concept of wide sense efficiency which reduces to the usual strict sense efficiency whenever the limiting distribution exists and is normal.

Examples of estimators with asymptotic variances which never exceed the inverse of the Fisher information and at some parameter points are below it were provided by Hodges and reported by LeCam [24]. Such estimators are called super-efficient (for example, see Zacks [53] p.208). LeCam [24] has shown that the set of parameter points,on which a super-efficient estimator has an asymptotic variance smaller than the inverse of the Fisher information, is of Lebesque measure zero. Rao [37] investigated the problems of asymptotic efficiency in a series of papers. He showed that super-efficient estimators, among consistent asymptotically normal ones, can exist if the approach to asymptotic normal distribution is not uniform on compact sets of the parameter points. Such a result was also established by Bahadur [3]. Thus, Rao suggested that one should consider only those consistent asymptotically normal estimators whose convergence in distribution to normal is uniform. He has shown that, under certain regularity conditions, the asymptotic variance of such estimators is always at least as large as the inverse of Fisher information. Thus, the uniformly best asymptotically normal estimators are the most efficient in this restricted class of estimators. Such estimators are called first-order efficient estimators by Rao.

The asymptotic properties of maximum likelihood estimator have been studied by many people under a variety of conditions. Cramér [14] gave proofs of weak consistency, asymptotic normality and asymptotic efficiency of MLE in the independent case, under some regularity conditions. These regularity conditions include the existence of the first three derivatives of the log-likelihood function and some uniform integrability conditions. An outline of the proof of strong consistency of the MLE was given by Wald [51] and Wolfowitz [52], but on the assumption that the estimate really maximises the likelihood function. Huber [19] established these results for the families of densities which do not satisfy the usual differentiability assumptions. Prakasa Rao [34] showed that MLE is strongly consistent and hyper-efficient for a special class of densities which do not satisfy the usual regularity conditions of Cramér [14]. Conditions under which MLE is uniformly consistent were presented recently by Moran [29, 30]. Ibragimov and Has'minskii [23] proved the asymptotic normality of the MLE without the assumptions of continuity of log-likelihood function and existence of the second derivative of the likelihood function. In all these cases, it is assumed that the observations, on which the MLEs are based, are independent and identically distributed.

These results have been extended in many directions. For example, some results were obtained for models in which the observations are independent but not identically distributed. It is worth mentioning that, under this situation, even if the observations are (non identically) normally distributed, the MLE may not be consistent and yet a consistent estimator does exist (see Zacks [53], p.236). Chao [12] established the strong consistency of MLE when the observations are independent but not identically distributed. In the same situation, Hoadley [18] gave conditions under which MLE is weakly consistent and asymptotically normal. His approach to consistency is similar to that taken by Wald [51].

Billingsley [8] proved weak consistency and asymptotic normality of an MLE for observations from a Markov process with stationary transition measures. His assumptions are analogous to those imposed by Cramér [14] in the case of independent random variables. The proof of asymptotic normality is based on a central limit theorem for martingales due to Billingsley [7]. For the same case, Roussas [42, 43] obtained strong consistency and asymptotic normality of the MLE, under different conditions. Some results are also obtained by Prakasa Rao [35] generalizing the results of Huber [19]. Sarma [45] proved that an MLE is also first-order efficient in addition to the properties of consistency and asymptotic normality.

The asymptotic properties of maximum likelihood estimators for the case of arbitrary stochastic process were first studied by Wald [50]. He gave conditions under which MLE is weakly consistent and proved that a consistent estimator is asymptotically efficient at least in the wide sense. M.M. Rao [39] extended these results to non-stable processes. Using a version of the central limit theorem due to P.Lévy, Silvey [48] pointed out the conditions under which an MLE is asymptotically normal. He gave conditions for weak consistency, but they were not expressed in terms of the probability distribution of the observations, and therefore, do not appear to be readily applicable. In a more recent work, Bar-Shalom [4] presented a set of conditions which ensure weak consistency and asymptotic efficiency of MLE in the case of dependent random variables.

Under a set of conditions which is stronger than that for consistency of MLE, LeCam [24, 26] proved the consistency of Bayes estimator for every parameter point in the independent case. Lindley [27] has established heuristically that MLE and Bayes estimators are asymptotically equivalent. LeCam [25] proved one of the fundamental results in the asymptotic theory of inference viz., the Bernstein-vonMises theorem regarding the convergence of posterior density to normal. Special cases of this theorem were first obtained by S. Bernstein and R. vonMises (for reference see LeCam[25]). Bickel and Yahav [6] and Chao [13] discussed the asymptotic behaviour of Bayes estimators. Recently, Ibragimov and Has'minskii [21, 22] studied the limiting properties of

Bayes estimators and maximum likelihood estimators under a different set of assumptions. Borwanker, Kallianpur and Prakasa Rao [11] extended the results of Bickel and Yahav[6] and LeCam [24, 26] to the Markov case.

In this thesis the asymptotic properties of maximum likelihood estimators and results concerning Bayes estimators are obtained in the case of dependent random variables. This is done in three different cases, namely, when the observations are from (i) stationary m-dependent sequences of random variables, (ii) stationary ϕ-mixing processes, and (iii) arbitrary stochastic processes.

Conditions are given which ensure weak consistency and first-order efficiency of MLE for stationary m-dependent random variables, in Chapter II. The approach adopted is the method of Bernstein, which he used to prove central limit theorem for m-dependent random variables. We were unable to give conditions for asymptotic normality in this case. But some comments are made about the asymptotic distribution. An example satisfying the regularity conditions in which MLE is asymptotically normal is given.

In Chapter III, weak consistency, asymptotic normality and first-order efficiency of approximate maximum likelihood estimator (AMLE) (for definition, see Section 3.5) are obtained for stationary ϕ-mixing processes. An alternative approach based on conditioning on past observations is used in this chapter. Under the set of conditions presented in this chapter, the proofs of weak consistency and asymptotic normality of AMLE depend on weak law of large numbers and central limit theorem for double sequences of random variables. Weak law of large numbers for double sequences of random variables is proved using Khintchine's method. This method was also used by Rosén [40] to prove central limit theorem. A particular case of the central limit theorem of Rosen [40] for double sequences of stationary ϕ-mixing random variables is used to prove the asymptotic normality of AMLE. Examples satisfying the conditions are given.

Strong consistency, asymptotic normality and first-order efficiency of MLE are proved in Chapter IV for arbitrary stochastic processes. The regularity conditions are expressed in terms of probability distribution of observations conditioned upon all past observations, as in Bar-Shalom [4]. In this case, we have also obtained Bernstein-vonMises theorem. An example is provided.

Cramér's approach is used in all the above results. Alternative possible approaches are the method of contiguity as given in detail in Roussas [44] or the method of convergence in distribution of stochastic processes as used in Prakasa Rao [34] and Ibragimov-Has'minskii [23].

Chapter V includes some comments on our work and suggestions for further research in this important field.

Whenever we need to refer to an equation C or a theorem C or a lemma C, in Section B of Chapter A, we do so by writing (A.B.C.) or theorem A.B.C. or lemma A.B.C., respectively. Throughout the following chapters, we write $(p)\lim_{n\to\infty}$, $(a.s)\lim_{n\to\infty}$ and $(\mathcal{D})\lim_{n\to\infty}$ to denote that a sequence of random variables converges in probability, converges almost surely and converges in distribution respectively.

CHAPTER II

MAXIMUM LIKELIHOOD ESTIMATION FOR STATIONARY m-DEPENDENT SEQUENCES

2.1 Introduction

Successive contributions by many authors, including Cramér [14], Wald [51], Huber [19], Prakasa Rao [34] and Moran [29,30] have yielded sufficiently weak regularity conditions under which the consistency, asymptotic normality and asymptotic efficiency of a maximum likelihood estimator (MLE) hold for independent and identically distributed random variables. These properties of MLEs are also studied for independent but not identically distributed random variables (see Chao [12] and Hoadley [18]) and for stationary Markov processes (see Billingsley [8], Sarma [45], Roussas [42,43] and Prakasa Rao [35]). The main tools used to prove consistency and asymptotic normality in these cases are laws of large numbers and central limit theorems respectively, for sums of random variables. So it is obvious that these results will extend to more general stochastic processes for which laws of large numbers and central limit theorems hold. Results of this type are available under various conditions on the type of dependence between the random variables (see, for example Ibragimov [20], Rosén [40] and Serfling [46]).

$t_1 < t_2 < \ldots < t_s$ is stochastically independent of any other subset$\{ X_{t_k}, X_{t_{k+1}}, \ldots\}$, $t_s < t_k < t_{k+1} < \ldots$, whenever $(t_k - t_s) > m$.

Let $\{X_n, n \geq 1\}$ be a stationary sequence of random variables and $\mathcal{B}$ be the σ-field generated by this sequence. Let T be the shift transformation corresponding to the stationary process $\{X_n, n \geq 1\}$ (for details of these concepts, we refer to Doob [16], p. 452). Then we have the following lemmas:

Lemma 2.2.1: A necessary and sufficient condition for a stationary process $\{X_n, n \geq 1\}$ to be ergodic is that for any two sets A and B measurable with respect to $\mathcal{B}$,

$$\lim_{n\to\infty} n^{-1} \sum_{j=o}^{n-1} P(B \cap T^{-j} A) = P(A)\, P(B). \tag{2.2.1}$$

Proof: For proof of this lemma, we refer to Rosenblatt [41], p. 110.

Remark 2.2.1: Equation (2.2.1) is implied by a stronger condition, namely

$$\lim_{j\to\infty} P(B \cap T^{-j} A) = P(A)\, P(B) \tag{2.2.2}$$

for any two sets measurable with respect to $\mathcal{B}$. This condition is commonly referred to as mixing condition and is, essentially, a form of asymptotic independence.

Lemma 2.2.2: Let $\mathcal{B}_0$ be a field generating the σ-field $\mathcal{B}$. If (2.2.2) holds for any pair of sets A and B in $\mathcal{B}_0$, then it holds for any pair in $\mathcal{B}$.

Proof: For proof, see Theorem 1.2 of Billingsley [9], p.12.

Lemma 2.2.3: Let $\{X_n, n \geq 1\}$ be a sequence of stationary m-dependent random variables. Then it is ergodic.

Proof: Let $\mathcal{B}$ be the σ-field generated by the sequence $\{X_n, n \geq 1\}$ and T be the shift transformation corresponding to this process. Let $\mathcal{B}_n$ be the field generated by the random variables $X_1, \ldots, X_n$. Then $\mathcal{B}$ is the σ-field generated by the field $\mathcal{B}_0 = \bigcup_{n=1}^{\infty} \mathcal{B}_n$. The lemma is proved, if we show that (2.2.2) is satisfied for any pair of sets A and B measurable with respect to $\mathcal{B}$. However, in view of Lemma 2.2.2, it is enough to verify (2.2.2) for any pair of sets A and B in $\mathcal{B}_n$, $n \geq 1$. Let A and B be the sets

$$A = \{\omega : X_1(\omega) \leq a_1, \ldots, X_n(\omega) \leq a_n\} , \quad (2.2.3)$$

$$B = \{\omega : X_1(\omega) \leq b_1, \ldots, X_n(\omega) \leq b_n\} . \quad (2.2.4)$$

where a_i and b_i, $1 \leq i \leq n$ are real numbers. Then

$$T^{-j} A = \{\omega : X_{1+j}(\omega) \leq a_1, \ldots, X_{n+j}(\omega) \leq a_n\} . \quad (2.2.5)$$

For $j > n + m - 1$, the sets B and $T^{-j}A$ are independent because of m-dependence of the sequence $\{X_n, n \geq 1\}$ of random variables. Further $P(T^{-j} A) = P(A)$, since the sequence is stationary. Therefore, for $j > n + m - 1$,

$$P(B \cap T^{-j} A) = P(B)\ P(A) \qquad (2.2.6)$$

so that (2.2.2) is satisfied for A and B. Hence the lemma follows.

Definition 2.2.2: A sequence of random variables $\{X_n, n \geq 1\}$ is said to be bounded in probability if, for any $\varepsilon > 0$, there exists an M_ε such that

$$\limsup_{n \to \infty} P[|X_n| \geq M_\varepsilon] \leq \varepsilon. \qquad (2.2.7)$$

The following lemma is well-known.

Lemma 2.2.4: Let $\{X_n, n \geq 1\}$ and $\{Y_n, n \geq 1\}$ be two arbitrary sequences of random variables. Then the following results are true.

(i) If $(p)\lim_{n\to\infty} X_n = X$ and f is a continuous function on R, then $(p)\lim_{n\to\infty} f(X_n) = f(X)$.

(ii) If $(p)\lim_{n\to\infty} X_n = 0$ and $\{Y_n\}$ is bounded in probability, then $\{X_n + Y_n\}$ is bounded in probability.

(iii) If $(p)\lim_{n\to\infty} X_n = 0$ and $\{Y_n\}$ is bounded in probability, then $(p)\lim_{n\to\infty} X_nY_n = 0$.

(iv) If $(p)\lim_{n\to\infty} X_n = c(\neq 0)$ and $\{Y_n\}$ is bounded in probability, then $\{Y_n/X_n\}$ is bounded in probability.

(v) If $(p)\lim_{n\to\infty} (X_n - Y_n) = 0$ and X_n converges to a random variable X in distribution or in probability, then Y_n also converges in distribution to X.

Proof: The proof of (i) can be found in Tucker [49] p.104. For the proofs of remaining results see M.M.Rao [38] and the references there.

Lemma 2.2.5: A sequence of jointly distributed random variables $\{X_n\}$ with zero mean and uniformly bounded variances obeys the strong law of large numbers, in the sense that $P[\lim_{n\to\infty} Z_n = 0] = 1$, if positive constants M and q exist such that for all integers n,

$$|E(X_n\ Z_n)| \le M\, n^{-q}\ , \tag{2.2.8}$$

where $$Z_n = n^{-1} \sum_{i=1}^{n} X_i .$$

Proof: For proof, we refer to Theorem 2B, Parzen [32] p. 420.

Lemma 2.2.6: Let $\{X_n^k,\ n = 1, 2, \ldots, k;\ k \ge 1\}$ be a double sequence of random variables satisfying the following conditions:

(i) The random variables X_n^k are m-dependent for any fixed $k \ge 1$,

(ii) $E(X_n^k) = 0$ for all n and k,

(iii) $\mathrm{Var}(S_k) = 1$, $k \ge 1$, where $S_k = \sum_{n=1}^{k} X_n^k$,

(iv) $\limsup_{k\to\infty} \sum_{n=1}^{k} \mathrm{Var}(X_n^k) < \infty$,

(v) $\lim_{k\to\infty} \sum_{n=1}^{k} \int_{|x|>\varepsilon} x^2\, dF_{kn}(x) = 0$

for every $\varepsilon > 0$ where $F_{kn}(x)$ is the distribution function of X_n^k. Then S_k converges in distribution to normal with mean zero and variance one.

Proof: For proof, see Theorem 13.1 , Rosén [40].

2.3 Regularity Conditions:

Let $\{X_n, n \geq 1\}$ be a stationary m-dependent sequence of random variables. Let $p(x_1, \ldots, x_n; \theta)$ denote the n-dimensional joint density function which depends on a single unknown parameter θ. The problem is to study the asymptotic properties of an MLE $\hat{\theta}_n$ of θ.

We denote $p(x_1, \ldots, x_n; \theta)$, $\phi_k(x_1, \ldots, x_n; \theta)$ and $p(x_{(i-1)k+r+1}, \ldots, x_{ik+r}; \theta)$ by $p_n(\theta)$, $\phi_k(\theta)$ and $L_{i,k}(\theta)$ respectively in this chapter. We write $P\{\cdot\}$ and $E\{\cdot\}$ instead of $P_{\theta_o}\{\cdot\}$ and $E_{\theta_o}\{\cdot\}$ respectively.

Definition 2.3.1: An estimator $\hat{\theta}_n \equiv \hat{\theta}_n(x_1, \ldots, x_n)$ of the unknown parameter θ, which maximises $\log p(x_1, \ldots, x_n; \theta)$ in a neighbourhood of the true parameter value θ_o, is said to be a maximum likelihood estimator (MLE) of θ.

Suppose the following regularity conditions are satisfied by the n-dimensional joint density function $p_n(\theta)$.

(H1) The parameter space Θ is an open interval in the real line and the true parameter value θ_o is an interior point of Θ.

(H2) $\frac{\partial}{\partial\theta} \log p_n(\theta)$ and $\frac{\partial^2}{\partial\theta^2} \log p_n(\theta)$ exist and are continuous almost surely for all $\theta \varepsilon \Theta$ and for all n.

(H3) Suppose that for all integers $k > 2m$, whenever n is of the form $\nu k+r$,

$$p_n(\theta) = p_r(\theta) \{ \prod_{i=1}^{\nu} L_{i,k}(\theta) \} \phi_k(\theta) \qquad (2.3.1)$$

(this representation is useful in view of m-dependence of the sequence of random variables $\{X_n, n \geq 1\}$) where $p_r(\theta)$ and $\phi_k(\theta)$ satisfy the following conditions for all $\theta \varepsilon \Theta$.

(i) $$\underset{k\to\infty}{\text{(p)lim}}\ \underset{n\to\infty}{\text{(p)lim}}\ n^{-1} \frac{\partial^2}{\partial\theta^2} \log p_r(\theta) = 0, \qquad (2.3.2)$$

(ii) $$\underset{k\to\infty}{\text{(p)lim}}\ \underset{n\to\infty}{\text{(p)lim}}\ n^{-1} \frac{\partial}{\partial\theta} \log \phi_k(\theta) = 0, \qquad (2.3.3)$$

and

(iii) for each $\theta \varepsilon \Theta$, there exists a neighbourhood $V_1(\theta)$ of θ such that

$$\underset{k\to\infty}{\text{(p)lim}}\ \underset{n\to\infty}{\text{(p)lim}}\ [\sup_{\theta' \varepsilon V_1(\theta)} \{ n^{-1} \frac{\partial^2}{\partial\theta^2} \log \phi_k(\theta') \}] = 0. \qquad \ldots(2.3.4)$$

(H4) For all $\theta \varepsilon \Theta$,

$$E_\theta \{\frac{\partial}{\partial\theta} \log p_n(\theta)\} = 0 \text{, for all } n, \qquad (2.3.5)$$

$$E_\theta \{\frac{\partial}{\partial\theta} \log p_n(\theta)\}^2 = - E_\theta\{\frac{\partial^2}{\partial\theta^2} \log p_n(\theta)\}$$

$$= I_n(\theta) < \infty \text{, for all } n, \qquad (2.3.6)$$

and

$$\lim_{n\to\infty} n^{-1} I_n(\theta) = i(\theta) \qquad (2.3.7)$$

where $i(\theta)$ is finite and positive.

(H5) For every $\theta \varepsilon \Theta$, there exists a neighbourhood $V(\theta)$ of θ such that, for every $\theta' \varepsilon V(\theta)$

$$\left|\frac{\partial^2}{\partial\theta^2} \log p_n(\theta) - \frac{\partial^2}{\partial\theta^2} \log p_n(\theta')\right| \le |\theta-\theta'|\ G(x_1,\dots,x_n)$$

where $G(x_1,\dots,x_n)$ is non-negative with $E_\theta[G(x_1,\dots,x_n)] = M_n(\theta)$, finite. Further,

$$\lim_{n\to\infty} n^{-1} M_n(\theta) = M(\theta) < \infty. \qquad (2.3.8)$$

Hereafter, we denote $\log p_n(\theta)$ by $\psi_n(\theta)$, $\log \phi_k(\theta)$ by $h_k(\theta)$ and $\log L_{i,k}(\theta)$ by $g_{i,k}(\theta)$. If $f(\theta)$ is a function differentiable twice with respect to θ, we denote the first and second derivatives of $f(\theta)$ by $f'(\theta)$ and $f''(\theta)$ respectively. Unless otherwise specified all statements which involve convergence of random variables in the following section are valid with respect to the true probability measure.

2.4 The Asymptotic Properties of a MLE:

We prove in this section the consistency and first-order efficiency of a MLE.

Theorem 2.4.1: Under the assumptions of Section 2.3, the likelihood equation has a root which is Weakly consistent.

Proof: Expanding $\psi_n'(\theta) = \frac{\partial}{\partial\theta} \log p_n(\theta)$ about the true value θ_o of the parameter θ, we obtain that

$$\psi_n'(\theta) = \psi_n'(\theta_o) + (\theta-\theta_o)\ \psi_n''(\theta_o)$$

$$+ (\theta-\theta_o)\ \{\psi_n''(\theta') - \psi_n''(\theta_o)\} \qquad (2.4.1)$$

where $\theta' = \theta_o + \alpha(\theta-\theta_o)$ and $|\alpha| < 1$. By (2.3.1) the right side can be written in the form

$$\{\psi_r'(\theta_o) + h_k'(\theta_o) + \sum_{i=1}^{\nu} g_{i,k}'(\theta_o)\}$$

$$+ (\theta-\theta_o)\ \{\psi_r''(\theta_o) + h_k''(\theta_o) + \sum_{i=1}^{\nu} g_{i,k}''(\theta_o)\}$$

$$+ (\theta-\theta_o)\ [\{\psi_r''(\theta') - \psi_r''(\theta_o)\} + \{h_k''(\theta') - h_k''(\theta_o)\}$$

$$+ \sum_{i=1}^{\nu}\{g_{i,k}''(\theta') - g_{i,k}''(\theta_o)\}].$$

It follows from (H5) that there exists a neighbourhood $V(\theta_o)$ of θ_o such that for $\theta' \in V(\theta_o)$

$$|\psi_r''(\theta') - \psi_r''(\theta_o)| \leq |\alpha(\theta-\theta_o)|\ G(x_1,\ldots,x_r)$$

and

$$|g_{i,k}''(\theta') - g_{i,k}''(\theta_o)| \leq |\alpha\ (\theta-\theta_o)|\ G(x_{(i-1)k+r+1},\ldots,x_{ik+r})$$

where $G(x_1,\ldots,x_r)$ and $G(x_{(i-1)k+r+1}, \ldots, x_{ik+r})$ are non-negative, $E[G(X_1,\ldots, X_r)] = M_r(\theta_o) < \infty$, $E[G(X_1, \ldots, X_k)] = M_k(\theta_o) < \infty$, and

$$\lim_{k\to\infty} k^{-1}\ E[G(X_1, \ldots, X_k)] = M(\theta_o) < \infty.$$

Hence we get that,

$$n^{-1}\ \psi_n'(\theta) = B_{o,k}^{(n)} + (\theta-\theta_o)\ B_{1,k}^{(n)} + (\theta-\theta_0)^2\alpha_n B_{2,k}^{(n)} \qquad (2.4.2)$$

where $|\alpha_n| < 1$ for all n, and

$$B_{o,k}^{(n)} = n^{-1} \{\psi_r'(\theta_o) + h_k'(\theta_o) + \sum_{i=1}^{\nu} g'_{i,k}(\theta_o)\}$$

$$+ (\theta-\theta_o)\, n^{-1} \{h_k''(\theta') - h_k''(\theta_o)\}, \tag{2.4.3}$$

$$B_{1,k}^{(n)} = n^{-1} \{\psi_r''(\theta_o) + h_k''(\theta_o) + \sum_{i=1}^{\nu} g_{i,k}''(\theta_o)\}$$

$$= n^{-1} \psi_n''(\theta_o) \tag{2.4.4}$$

and

$$B_{2,k}^{(n)} = n^{-1} \{G(x_1, \ldots, x_r)$$

$$+ \sum_{i=1}^{\nu} G(x_{(i-1)k+r+1}, \ldots, x_{ik+r})\}. \tag{2.4.5}$$

Since $\{X_n, n \geq 1\}$ is a stationary m-dependent sequence of random variables, for any fixed $k \geq 2m$ and any fixed r such that $o \leq r < k$, $\{g_{i,k}'(\theta_o), i \geq 1\}$, $\{g_{i,k}''(\theta_o), i \geq 1\}$ and $\{G(x_{(i-1)k+r+1}, \ldots, x_{ik+r}), i \geq 1\}$ are stationary 1-dependent sequences of random variables and hence, by Lemma 2.2.3, are ergodic. Furthermore, by (H4) and (H5), it is clear that

$$E[g_{1,k}'(\theta_o)] = 0,$$

$$E[g_{1,k}''(\theta_o)] = -I_k(\theta_o) < \infty,$$

and

$$E[G(X_1, \ldots, X_k)] = M_k(\theta_o) < \infty.$$

Therefore, by ergodic theorem, we get

$$(a.s)\lim_{\nu\to\infty} \nu^{-1} \sum_{i=1}^{\nu} g'_{i,k}(\theta_o) = E[g'_{1,k}(\theta_o)] = 0,$$

$$(a.s)\lim_{\nu\to\infty} \nu^{-1} \sum_{i=1}^{\nu} g''_{i,k}(\theta_o) = E[g''_{1,k}(\theta_o)] = - I_k(\theta_o),$$

and

$$\begin{aligned}(a.s)\lim_{\nu\to\infty} \nu^{-1} \sum_{i=1}^{\nu} G(X_{(i-1)k+r+1}, \ldots, X_{ik+r}) &= E[G(X_{r+1},\ldots,X_{r+k})] \\ &= E[G(X_1, \ldots, X_k)] \\ &= M_k(\theta_o).\end{aligned}$$

Since $\nu \to \infty$ as $n \to \infty$, we have

$$\begin{aligned}(a.s)\lim_{n\to\infty} n^{-1} \sum_{i=1}^{\nu} g'_{i,k}(\theta_o) &= k^{-1} E[g'_{1,k}(\theta_o)] \\ &= 0, \qquad (2.4.6)\end{aligned}$$

$$(a.s)\lim_{n\to\infty} n^{-1} \sum_{i=1}^{\nu} g''_{i,k}(\theta_o) = k^{-1} I_k(\theta_o), \qquad (2.4.7)$$

and

$$\begin{aligned}(a.s)\lim_{n\to\infty} n^{-1} \sum_{i=1}^{\nu} G(X_{(i-1)k+r+1}, \ldots, X_{ik+r}) &= k^{-1} M_k(\theta_o). \qquad (2.4.8)\end{aligned}$$

Therefore, it follows from assumptions (H4) and (H5) that

$$\lim_{k\to\infty} (a.s)\lim_{n\to\infty} n^{-1} \sum_{i=1}^{\nu} g'_{i,k}(\theta_o) = 0, \qquad (2.4.9)$$

$$\lim_{k\to\infty} (a.s)\lim_{n\to\infty} n^{-1} \sum_{i=1}^{\nu} g''_{i,k}(\theta_o) = - i_o, \qquad (2.4.10)$$

and

$$\lim_{k\to\infty} (a.s) \lim_{n\to\infty} n^{-1} \sum_{i=1}^{\nu} G(X_{(i-1)k+r+1}, \ldots, X_{ik+r}) = M_o$$

...(2.4.11)

where $i_o = i(\theta_o)$ and $M_o = M(\theta_o)$. Now, for any $\varepsilon > 0$,

$$P[n^{-1}|\psi_r'(\theta_o)| > \varepsilon] \le E[\psi_r'(\theta_o)]^2 (n\varepsilon)^{-2}$$

$$= I_r(\theta_o) (n\varepsilon)^{-2}.$$

Since $I_r(\theta_o)$ is finite, the right side converges to zero as $n \to \infty$. Hence we obtain that

$$\lim_{k\to\infty} (p) \lim_{n\to\infty} n^{-1} \psi_r'(\theta_o) = 0. \quad (2.4.12)$$

By (2.3.3)

$$(p) \lim_{k\to\infty} (p) \lim_{n\to\infty} n^{-1} h_k'(\theta_o) = 0, \quad (2.4.13)$$

and by (2.3.4), there exists a neighbourhood $V_1(\theta_o)$ of θ_o such that

$$(p) \lim_{k\to\infty} (p) \lim_{n\to\infty} [\sup_{\theta' \varepsilon V_1(\theta_o)} n^{-1} h_k''(\theta')] = 0.$$

Therefore

$$(p) \lim_{k\to\infty} (p) \lim_{n\to\infty} n^{-1} \{h_k''(\theta') - h_k''(\theta_o)\} = 0. \quad (2.4.14)$$

From (2.4.9), (2.4.12) - (2.4.14), it follows that

$$(p) \lim_{k\to\infty} (p) \lim_{n\to\infty} B_{o,k}^{(n)} = 0. \quad (2.4.15)$$

By (2.3.2), we have

$$(p) \lim_{k\to\infty} (p) \lim_{n\to\infty} n^{-1} \psi_r''(\theta_o) = 0, \quad (2.4.16)$$

and by (2.3.4),

$$\underset{k\to\infty}{(p)\lim}\ \underset{n\to\infty}{(p)\lim}\ n^{-1}\ h_k''(\theta_o) = 0. \tag{2.4.17}$$

Hence from (2.4.10), (2.4.16) and (2.4.17) we get

$$\underset{k\to\infty}{(p)\lim}\ \underset{n\to\infty}{(p)\lim}\ B_{1,k}^{(n)} = \underset{n\to\infty}{(p)\lim}\ n^{-1}\ \psi_n''(\theta_o)$$

$$= -\ i_o. \tag{2.4.18}$$

By Tchebyshev's inequality, for every $\varepsilon > 0$,

$$P[n^{-1}\ G(X_1,\ \ldots,\ X_r) > \varepsilon] \le (n\varepsilon)^{-1}\ E[G(X_1,\ldots,X_r)]$$

$$= (n\varepsilon)^{-1}\ M_r(\theta_o).$$

Since $M_r(\theta_o)$ is finite, the right side tends to zero as $n \to \infty$. Hence

$$\lim_{k\to\infty}\ \underset{n\to\infty}{(p)\lim}\ n^{-1}\ G(X_1,\ldots,X_r) = 0. \tag{2.4.19}$$

Therefore from (2.4.5), (2.4.11) and (2.4.19) we get that

$$\lim_{k\to\infty}\ \underset{n\to\infty}{(p)\lim}\ B_{2,k}^{(n)} = M_o. \tag{2.4.20}$$

Choose $\delta > 0$ such that $\{\theta : \theta_o - \delta \le \theta \le \theta_o + \delta\}$ is contained in $V(\theta_o) \cap V_1(\theta_o)$. Let

$$\underset{n\to\infty}{(p)\lim}\ B_{o,k}^{(n)} = B_{o,k}\ , \tag{2.4.21}$$

$$\underset{n\to\infty}{(p)\lim}\ B_{1,k}^{(n)} = B_{1,k}\ , \tag{2.2.22}$$

and

$$\underset{n\to\infty}{(p)\lim}\ B_{2,k}^{(n)} = B_{2,k}. \tag{2.4.23}$$

Here

$$B_{o,k} = 0 \text{ almost surely for every } k \geq 2m,$$

$$(p)\lim_{k\to\infty} B_{1,k} = - i_o,$$

and

$$B_{2,k} = \frac{M_k(\theta_o)}{k} \text{ almost surely for every } k \geq 2m.$$

Note that $\{B_{o,k}\}$ and $\{B_{2,k}\}$ are constant valued random variables with probability one. Let $\varepsilon > 0$. Then there exists a positive integer $k_o = k_o(\varepsilon)$ such that, for all $k \geq k_o(\varepsilon)$,

$$P[|B_{1,k} + i_o| > \delta/2] \leq \varepsilon/6,$$

and

$$|B_{2,k}| < (M_o+1)/2,$$

with probability one. From (2.4.21)-(2.4.23) we get that for $n \geq n_o(k_o,\varepsilon)$,

$$P[|B^{(n)}_{o,k_o}| > \delta^2] \leq \varepsilon/3,$$

$$P[|B^{(n)}_{1,k_o} - B_{1,k_o}| > \delta/2] \leq \varepsilon/6,$$

and

$$P[|B^{(n)}_{2,k_o} - B_{2,k_o}| > (M_o+1)/2] \leq \varepsilon/3.$$

Let $S_1 = \{|B^{(n)}_{o,k_o}| > \delta^2\}$,

$$S_2 = \{|B^{(n)}_{1,k_o} + i_o| > \delta\},$$

and

$$S_3 = \{|B^{(n)}_{2,k_o}| > M_o + 1\}.$$

Therefore, from (2.4.15), (2.4.18) and (2.4.20) it follows that for every $n \geq n_o(k_o,\varepsilon)$, $P(S_i) \leq \varepsilon/3$ for $i = 1,2,3$. Hence $P(S) \leq \varepsilon$ for $n \geq n_o(k_o,\varepsilon)$, where $S = \bigcup_{i=1}^{3} S_i$. For $\theta = \theta_o \pm \delta$, (2.4.2) can be written as

$$n^{-1}\,\psi_n'(\theta) = B^{(n)}_{o,k_o} \pm B^{(n)}_{1,k_o}\,\delta + B^{(n)}_{2,k_o}\,\alpha_n\,\delta^2. \tag{2.4.24}$$

For $n > n_o(k_o,\varepsilon)$,

$$P[|B^{(n)}_{o,k_o} + \delta^2\,\alpha_n\,B^{(n)}_{2,k_o}| < \delta^2(M_o+2);\ -\delta\,B^{(n)}_{1,k_o} \geq -\delta(i_o-\delta)]$$

$$> 1 - \varepsilon.$$

So if we choose $\delta < i_o(M_o+3)^{-1}$, the sign of

$$-B^{(n)}_{o,k_o} \pm B^{(n)}_{1,k_o}\delta + B^{(n)}_{2,k_o}\,\alpha_n\,\delta^2$$

depends on that of $-\delta B^{(n)}_{1,k_o}$ with probability exceeding $1-\varepsilon$ for $n > n_o(k_o,\varepsilon)$. Therefore, for every $n \geq n_o(k_o,\varepsilon)$,

$$\psi_n'(\theta) > 0 \quad \text{for } \theta = \theta_o - \delta$$

and

$$\psi_n'(\theta) < 0 \quad \text{for } \theta = \theta_o + \delta$$

with probability exceeding $1-\varepsilon$. Since $\psi_n'(\theta)$ is continuous, there exists a value $\hat{\theta}_n$ in $[\theta_o - \delta,\ \theta_o + \delta]$ of θ, with probability greater than $1-\varepsilon$ for $n > n_o(k_o,\varepsilon)$ such that

$$\psi_n'(\hat{\theta}_n) = \frac{\partial}{\partial\theta} \log p_n(\hat{\theta}_n) = 0.$$

This proves that $\hat{\theta}_n$ exists and it is weak consistent since $P(|\hat{\theta}_n - \theta_o| \leq \delta) > 1-\varepsilon$ for $n > n_o(k_o,\varepsilon)$. Hence the theorem follows.

Definition 2.4.1. An estimator T_n of θ is said to be asymptotically efficient of first-order, if there exists two constants α and $\beta(\neq o)$, not depending on the sample but may depend on θ, such that

$$|\{I_n(\theta)\}^{-1/2} \frac{\partial}{\partial\theta} \log p_n(\theta) - \alpha - \beta(T_n-\theta)| \qquad (2.4.25)$$

converges to zero in probability as $n \to \infty$.

Definition 2.4.2: A sequence $\{T_n\}$ of estimators of θ is said to be asymptotically efficient in the wide sense, if there exists another sequence of random variables $\{W_n\}$ such that

(i) $\lim\limits_{n\to\infty} E_\theta(W_n) = 0$, (2.4.26)

(ii) $\lim\limits_{n\to\infty} E_\theta(W_n^2) = 1$, (2.4.27)

and

(iii) $(p)\lim\limits_{n\to\infty} [\{I_n(\theta)\}^{1/2} (T_n-\theta) - W_n] = 0.$ (2.4.28)

Remark 2.4.1: Suppose that (2.4.25) is satisfied for a sequence of estimators $\{T_n\}$ of θ with $\alpha \equiv 0$ and $\beta(\theta) = \{I_n(\theta)\}^{1/2}$. Then (2.4.26) - (2.4.28) will be satisfied with $W_n = \frac{\partial}{\partial\theta} \log p_n(\theta)$. Hence if a sequence of estimators $\{T_n\}$ of θ is asymptotically efficient of first-order, then it is asymptotically efficient in wide sense.

Now, we prove the following theorem:

Theorem 2.4.2: Under assumptions of Section 2.3, a consistent maximum likelihood estimator is asymptotically efficient of first-order.

Proof: We have $I_n(\theta_o) \neq 0$ for large n since $i_o > 0$ by assumption (H4). Hence, if $\hat{\theta}_n$ is a consistent root of the likelihood equation, then, from (2.4.1), it follows that

$$\{I_n(\theta_o)\}^{-1} [\psi_n'(\theta_o) + (\hat{\theta}_n-\theta_o) \{\psi_n''(\theta_o) + U_n\}] = 0 \quad (2.4.29)$$

where

$$U_n = \psi_n''(\theta') - \psi_n''(\theta_o)$$

and $|\theta'-\theta_o| \leq |\hat{\theta}_n-\theta_o|$. Since $\hat{\theta}_n$ is a consistent estimator and $\psi_n''(\theta)$ is continuous, by (i) of Lemma 2.2.4, we get

$$(p)\lim_{n\to\infty} U_n = 0.$$

Further, (2.3.7) and (2.4.18) imply that

$$(p)\lim_{n\to\infty} \{I_n(\theta_o)\}^{-1} \psi_n''(\theta_o) = -1.$$

Therefore, by (2.4.29),

$$\{I_n(\theta_o)\}^{-1/2} \psi_n'(\theta_o) - (\hat{\theta}_n-\theta_o) \{1+o_p(1)\} \{I_n(\theta_o)\}^{1/2} + \xi_n = 0 \quad (2.4.30)$$

where $o_p(1)$ is a term which converges to zero in probability as $n \to \infty$ and

$$\xi_n = \{I_n(\theta_o)\}^{-1/2} (\hat{\theta}_n - \theta_o) U_n \tag{2.4.31}$$

which tends to zero in probability as $n \to \infty$. (2.4.31) can be written in the form

$$Y_n = Z_n (1 + o_p(1)) + o_p(1) \tag{2.4.32}$$

where

$$Y_n = \{I_n(\theta_o)\}^{-1/2} \psi_n'(\theta_o), \tag{2.4.33}$$

and

$$Z_n = \{I_n(\theta_o)\}^{1/2} (\hat{\theta}_n - \theta_o). \tag{2.4.34}$$

Since $E(Y_n) = 0$ and Var $Y_n = 1$, $\{Y_n\}$ is bounded in probability. Then, $\{Y_n - o_p(1)\}$ is bounded in probability by (ii) of Lemma 2.2.4 so that $\{Z_n\}$ is bounded in probability by (iv) of Lemma 2.2.4. Now an application of (iii) of Lemma 2.2.4 gives that

$$(p)\lim_{n\to\infty} Z_n\, o_p(1) = 0. \tag{2.4.35}$$

Therefore, from (2.4.30) - (2.4.35), it follows that

$$\left|\{I_n(\theta_o)\}^{-1/2} \psi_n'(\theta_o) - (\hat{\theta}_n - \theta_o) \{I_n(\theta_o)\}^{1/2}\right|$$

$$= |Z_n\, o_p(1) + \xi_n|, \tag{2.4.36}$$

which tends to zero in probability as $n \to \infty$. Hence the MLE $\hat{\theta}_n$ of θ is asymptotically efficient of first-order.

Remark 2.4.2: By (2.4.32) - (2.4.35), it follows that $\hat{\theta}_n$ is also asymptotically efficient in wide sense in view of Remark 2.4.1.

Remark 2.4.3: If the sequence $\{Y_n\}$ converges in distribution to normal, then the efficiency in wide sense coincides with the classical strict sense concept, that is with Fisher efficiency (for definition see Cramér [14]).

Under the set of assumptions given in Section 2.3, we are unable to prove the asymptotic normality of MLE. However, we can make the following remark about the asymptotic distribution of MLE.

Remark 2.4.4: Since $\hat{\theta}_n$ is a MLE, we have (2.4.29) - (2.4.35). Therefore,

$$(p)\lim_{n\to\infty} (Y_n - Z_n) = 0. \tag{2.4.37}$$

Hence if Y_n converges to Y in distribution as $n \to \infty$ then, by (v) of Lemma 2.2.4, it follows that Z_n also converges in distribution to Y. Consequently, in order to obtain the asymptotic distribution of Z_n, it suffices to find the limiting distribution of Y_n as $n \to \infty$, if it exists.

In general, if $p(x_1,\ldots,x_n;\theta)$ is of the form given in (2.3.1), it is difficult to give conditions under which Y_n converges to a normal distribution. M.M.Rao [39] proved that Y_n will have an asymptotic distribution if the correlation between $\psi_n'(\theta)$ and $\psi_{n+r}'(\theta)$ tends to one uniformly in r as $n \to \infty$. This condition is difficult to verify even in the m-dependent case. In some special cases, it can be shown that the asymptotic distribution of MLE is normal. One such example is given in the next section.

2.5 Example:

In this section, we give an example which satisfies the conditions of Section 2.3. Let the sequence $\{X_n, n \geq 1\}$ of random variables be defined by

$$X_n = Y_n - \rho\, Y_{n-1} \tag{2.5.1}$$

where $\{Y_n, n \geq 0\}$ is an independent sequence of random variables, each Y_n being $N(\mu, 1)$ where μ is unknown and $|\rho| < 1$ is a known quantity. We assume that $\{X_n\}$ is observed. This is the first-order moving average process. The problem is to obtain the asymptotic properties of an MLE of μ. We show that the conditions of Section 2.3 are satisfied and thus establish the weak consistency and first-order efficiency of the MLE $\hat{\mu}_n$ of μ. Further we prove that $\hat{\mu}_n$ is asymptotically normal. It is clear from (2.5.1) that

$$E(X_n) = \mu(1-\rho)\,, \tag{2.5.2}$$

$$\operatorname{Var}(X_n) = 1 + \rho^2\,, \tag{2.5.3}$$

$$\begin{aligned} \operatorname{Cov}(X_n, X_m) &= -\rho \quad \text{for } |n-m| = 1 \\ &= 0 \quad \text{for } |n-m| > 1 \end{aligned} \tag{2.5.4}$$

and the n-dimensional joint density function which depends on the single unknown parameter μ is given by $p_n(\mu)$

$$= (2\pi)^{n/2} (\det\Gamma_n)^{-1/2} \exp[-\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_n^{ij} (x_i - C\mu)(x_j - C\mu)] \quad \ldots (2.5.5)$$

where $C = 1-\rho$ and Γ_n is the variance-covariance matrix of $(X_1, \ldots, X_n)$. The elements of Γ_n are given by (2.5.3) and (2.5.4). γ_n^{ij} is the i,j th element of Γ_n^{-1}. Then we have (see Shaman [47]),

$$\gamma_n^{ij} = \rho^{i-j}(1-\rho^{2j})(1-\rho^{2n-2i+2})\{(1-\rho^2)(1-\rho^{2n+2})\}^{-1} \quad \ldots (2.5.6)$$

for $i \geq j$. From (2.5.3) - (2.5.5), it follows that the sequence $\{X_n, n \geq 1\}$ of random variables is stationary and 1-dependent. Now, we have

$$\frac{\partial}{\partial\mu} \log p_n(\mu) = C \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_n^{ij}(x_i + x_j - 2C\mu)$$

$$= 2C \sum_{i=1}^{n} \gamma_n^{i\cdot}(x_i - C\mu) \quad (2.5.7)$$

and

$$\frac{\partial^2}{\partial\mu^2} \log p_n(\mu) = -C^2 \times 2 \sum_{i=1}^{n} \gamma_n^{i\cdot} \quad (2.5.8)$$

where

$$\gamma_n^{i\cdot} = \sum_{j=1}^{n} \gamma_n^{ij} = \sum_{j=1}^{i-1} \gamma_n^{ij} + \sum_{j=i}^{n} \gamma_n^{ij}$$

$$= \sum_{j=1}^{i-1} \gamma_n^{ij} + \sum_{j=i}^{n} \gamma_n^{ji} .$$

We have

$$\sum_{j=1}^{i-1} \gamma_n^{ij} = q_n^{-1} \sum_{j=1}^{i-1} \{\rho^{i-j}(1-\rho^{2j})(1-\rho^{2n-2i+2})$$

by (2.5.6), where

$$q_n = (1-\rho^2)(1-\rho^{2n+2}) \quad . \tag{2.5.9}$$

Therefore, we get

$$\sum_{j=1}^{i-1} \gamma_n^{ij} = \{(1-\rho)q_n\}^{-1} \; [\rho-\rho^{i+1}-\rho^{2n-2i+3}+\rho^{2n-i+3}]\,(1-\rho^{i-1})$$

$$= \{(1-\rho)q_n\}^{-1} \; \rho(1-\rho^{i-1})(1-\rho^{i})(1-\rho^{2n-2i+2}) \; .$$

Similarly, by (2.5.6), we get

$$\sum_{j=i}^{n} \gamma_n^{ji} = \{(1-\rho)q_n\}^{-1} \; (1-\rho^{2i})(1-\rho^{n-i+1})(1-\rho^{n-i+2})$$

so that

$$\gamma_n^{i\cdot} = \{q_n \; (1-\rho)\}^{-1} \; [\rho(1-\rho^{i-1})(1-\rho^{i})(1-\rho^{2n-2i+2})$$

$$+ (1-\rho^{2i})(1-\rho^{n-i+1})(1-\rho^{n-i+2})]$$

$$= \{(1-\rho)q_n\}^{-1} \; (1-\rho^{i})(1-\rho^{n-i+1}) \; [\rho(1-\rho^{i-1})(1+\rho^{n-i+1})$$

$$+ (1+\rho^{i})(1-\rho^{n-i+2})]$$

$$= (1-\rho^{i})(1-\rho^{n-i+1}) \; t_n \tag{2.5.10}$$

where

$$t_n = (1-\rho^{n+1}) \; \{(1-\rho)^2(1-\rho^{2n+2})\}^{-1} . \tag{2.5.11}$$

Therefore,

$$\sum_{i=1}^{n} \gamma_n^{i\cdot} = t_n \sum_{i=1}^{n} (1-\rho^i)(1-\rho^{n-i+1})$$

$$= t_n[n(1+\rho^{n+1}) - 2\rho(1-\rho^n)(1-\rho)^{-1}]$$

$$= n(1-\rho)^{-2} - 2\rho t_n (1-\rho^n)(1-\rho)^{-1}.$$

Hence, by (2.5.8), we have

$$\frac{\partial^2}{\partial\mu^2} \log p_n(\mu) = -2[n-2\rho t_n (1-\rho^n)(1-\rho)]. \tag{2.5.12}$$

It can be seen from (2.5.7) - (2.5.12) that

$$E_\mu[\frac{\partial^2}{\partial\mu^2} \log p_n(\mu)] = - E_\mu[\frac{\partial}{\partial\mu} \log p_n(\mu)]^2$$

after long calculations. Further from (2.5.12) we get that for all μ

$$i(\mu) = \lim_{n\to\infty} n^{-1} E_\mu[\frac{\partial}{\partial\mu} \log p_n(\mu)]^2$$

$$= 2. \tag{2.5.13}$$

We shall now show that the conditions (H3) and (H5) are satisfied by this process. Clearly, since $o \leq r < k$, from (2.5.12) we get that

$$\lim_{n\to\infty} n^{-1} \frac{\partial^2}{\partial\mu^2} \log p_r(\mu) = 0$$

for fixed k, which in turn shows that

$$\lim_{k\to\infty} \lim_{n\to\infty} n^{-1} \frac{\partial^2}{\partial\mu^2} \log p_r(\mu) = 0 \tag{2.5.14}$$

Furthermore, we have

$$n^{-1} \frac{\partial}{\partial\mu} \log \phi_k(\mu) = n^{-1}\, C\, 2\sum_{i=1}^{n} \gamma_n^{i\cdot}(x_i - C\mu) - n^{-1} \frac{\partial}{\partial\mu} \log p_r(\mu)$$

$$- n^{-1} \sum_{i=1}^{\nu} \frac{\partial}{\partial\mu} \log p(x_{(i-1)k+r+1},\ldots,x_{ik+r}). \qquad \ldots(2.5.15)$$

Let

$$z_n = C\, n^{-1} \sum_{i=1}^{n} \gamma_n^{i\cdot}(X_i - C\mu).$$

Then

$$|E[C\gamma_n^{n\cdot}(X_n - C\mu)\ z_n]|$$

$$= n^{-1}\, C^2 |\gamma_n^{n\cdot}\ \gamma_n^{(n-1)\cdot}|\, |E\{(X_n - C\mu)(X_{n-1} - C\mu)\}|$$

$$+ n^{-1}\, C^2\, |\gamma_n^{n\cdot}|^2\ |E(X_n - C\mu)^2|.$$

From (2.5.10) and (2.5.11) we have

$$|\gamma_n^{i\cdot}| \leq 8\ \{(1-\rho)^3\ (1+\rho)\}^{-1} \qquad (2.5.16)$$

For all $i = 1,2,\ldots,n$. Therefore,

$$|E[C\gamma_n^{n\cdot}(X_n - C\mu)\ z_n]| \leq 64(|\rho| + 1 + \rho^2)\ \{n(1-\rho^4)(1+\rho)^2\}^{-1}$$

$$\leq 192\{n(1-\rho^4)(1+\rho)^2\}^{-1}.$$

Further

$$\mathrm{Var}(CX_n\ \gamma_n^{n\cdot}) = C^2(\gamma_n^{n\cdot})^2\ (1+\rho^2)$$

$$\leq 64\{(1-\rho)^4 (1+\rho)^2\}^{-1} (1+\rho^2)$$

$$\leq 128 \{(1-\rho)^4 (1+\rho)^2\}^{-1} ,$$

so that $\mathrm{Var}(CX_n \gamma_n^{n\cdot})$ are uniformly bounded in n. Hence, by Lemma 2.2.5, we get that

$$(a.s)\lim_{n\to\infty} C\, n^{-1} \sum_{i=1}^{n} \gamma_n^{i\cdot}(X_i - C\mu) = 0. \tag{2.5.17}$$

Now, for any $\varepsilon > 0$

$$P[n^{-1} |\frac{\partial}{\partial\mu} \log p_r(\mu)| > \varepsilon] \leq (n\varepsilon)^{-2} E_\mu[\frac{\partial}{\partial\mu} \log p_r(\mu)]^2$$

$$= (n\varepsilon)^{-2}[r-2\rho t_r (1-\rho^r)(1-\rho)]$$

by (2.5.12). Since $0 \leq r < k$, the right hand side tends to zero as $n \to \infty$. Therefore,

$$\lim_{k\to\infty} (p)\lim_{n\to\infty} n^{-1} \frac{\partial}{\partial\mu} \log p_r(\mu) = 0. \tag{2.5.18}$$

Since $\{ \frac{\partial}{\partial\mu} \log p(X_{(i-1)k+r+1}, \ldots, X_{ik+r}),\ i \geq 1\}$ is a stationary 1-dependent sequence of random variables, it is ergodic and hence, by ergodic theorem, the third term on the right hand side of (2.5.15) converges almost surely to zero as $n \to \infty$, for every fixed k. Hence by (2.5.15), (2.5.17) and (2.5.18) we get that

$$\lim_{k\to\infty} (p)\lim_{n\to\infty} n^{-1} \frac{\partial}{\partial\mu} \log \phi_k(\mu) = 0 , \tag{2.5.19}$$

which is (ii) of condition (H3). Now, consider

$$n^{-1} \frac{\partial^2}{\partial\mu^2} \log \phi_k(\mu) = -C^2\, n^{-1} 2 \sum_{i=1}^{n} \gamma_n^{i\cdot} + n^{-1} \frac{\partial^2}{\partial\mu^2} \log p_r(\mu)$$

$$+ n^{-1} \sum_{i=1}^{\nu} \frac{\partial^2}{\partial\mu^2} \log p(x_{(i-1)k+r+1}, \ldots, x_{ik+r}) \qquad \ldots(2.5.20)$$

$$= -C^2\, n^{-1} 2 \sum_{i=1}^{n} \gamma_n^{i\cdot} + n^{-1}\, C^2\, 2 \sum_{i=1}^{r} \gamma_r^{i\cdot}$$

$$+ C^2\, n^{-1} 2 \sum_{i=1}^{\nu} \sum_{j=1}^{k} \gamma_k^{j\cdot} \,.$$

$$= 2\rho\, \{n(1-\rho)\}^{-1} 2[(1-\rho^n)t_n - (1-\rho^r)t_r$$

$$-\nu(1-\rho^k)t_k] \qquad (2.5.21)$$

by (2.5.12). Since this expression is independent of μ,

$$\sup_{\mu' \in V_1(\mu)} n^{-1} \frac{\partial^2}{\partial\mu^2} \log \phi_k(\mu') = n^{-1} \frac{\partial^2}{\partial\mu^2} \log \phi_k(\mu)$$

for any neighbourhood $V_1(\mu)$ of μ. By (2.5.21), for fixed k, we have

$$\lim_{n\to\infty} n^{-1} \frac{\partial^2}{\partial\mu^2} \log \phi_k(\mu) = -2\rho\{k(1-\rho)\}^{-1} (1-\rho^k)t_k,$$

so that

$$\lim_{k\to\infty} \lim_{n\to\infty} n^{-1} \frac{\partial^2}{\partial\mu^2} \log \phi_k(\mu) = 0.$$

Hence

$$\lim_{k\to\infty} \lim_{n\to\infty} [n^{-1} \sup_{\mu' \in V_1(\mu)} \frac{\partial^2}{\partial\mu^2} \log \phi_k(\mu)] = 0, \qquad (2.5.22)$$

For any neighbourhood $V_1(\mu)$ of μ. Thus (2.3.4) is satisfied. Since $\frac{\partial^2}{\partial\mu^2}\log p_n(\mu)$ is independent of $X_1,\ldots,X_n$, (H5) will be satisfied with $G(x_1,\ldots,x_n) \equiv 0$. Thus all the conditions of Section 2.3 are satisfied and the MLE of μ is a consistent estimator by Theorem 2.4.1 and first-order efficient by Theorem 2.4.2.

Now, we shall prove the asymptotic normality of the MLE $\hat{\mu}_n$ of μ. Let μ_o be the true value of the parameter μ. Then

$$Y_n = \{I_n(\mu_o)\}^{-1/2} \frac{\partial}{\partial\mu} \log p_n(\mu_o). \qquad (2.5.23)$$

By (2.4.37), in order to find the asymptotic distribution of $Z_n = \{I_n(\mu_o)\}^{1/2} (\hat{\mu}_n-\mu_o)$, it is enough if we find the asymptotic distribution of Y_n. We have

$$Y_n = \{n/I_n(\mu_o)\}^{1/2}\, n^{-1/2} \frac{\partial}{\partial\mu} \log p_n(\mu_o). \qquad (2.5.24)$$

Consider

$$n^{-1/2} \frac{\partial}{\partial\mu} \log p_n(\mu_o) = C\, n^{-1/2}\, 2\sum_{i=1}^{n} \gamma_n^{i\cdot} (X_i-C\mu_o)$$

$$= C\cdot s_n\, (n\, s_n^2)^{-1/2}\, 2\sum_{i=1}^{n} \gamma_n^{i\cdot}(X_i-C\mu_o)$$

where

$$s_n^2 = 4C^2\, n^{-1}\, E[\sum_{i=1}^{n} \gamma_n^{i\cdot} (X_i-C\mu_o)]^2$$

$$= n^{-1}\, E[C \frac{\partial}{\partial\mu} \log p_n(\mu_o)]^2$$

$$= -n^{-1}\, E[\frac{\partial^2}{\partial\mu^2} \log p_n(\mu_o)]$$

$$= [2-2\rho t_n \; (1-\rho^n) \; \{n(1-\rho)\}^{-1}] \qquad (2.5.25)$$

by (2.5.12). Let

$$\xi_n = \sum_{i=1}^{n} Y_{i,n} ,$$

where

$$Y_{i,n} = C \, \gamma_n^{i\cdot} \, s_n^{-1} \, n^{-1/2} \, (X_i - C\mu_o) \; .$$

Then it is clear that,

(i) the random variables $Y_{i,n}$ are 1-dependent for fixed n,

(ii) $E(Y_{i,n}) = 0$ for all i and n , and

(iii) $\limsup_{n\to\infty} \sum_{i=1}^{n} E(Y_{i,n})^2$

$$= \limsup_{n\to\infty} C^2 s_n^{-2} \limsup_{n\to\infty} n^{-1} \sum_{i=1}^{n} (\gamma_n^{i\cdot})^2 \, E(X_i - C\mu_o)^2$$

$$= (1+\rho^2) \; (1-\rho)^2 \limsup_{n\to\infty} n^{-1} \sum_{i=1}^{n} (\gamma_n^{i\cdot})^2 \; .$$

Since $E(X_i - C\mu_o)^2 = 1+\rho^2$, $C^2 = (1-\rho)^2$ and $\lim_{n\to\infty} s_n^2 = 1$ by (2.5.25). Hence by (2.5.16),

$$\limsup_{n\to\infty} \sum_{i=1}^{n} E(Y_{i,n})^2 \le 64(1+\rho^2)(1-\rho)^2\{(1-\rho)^2(1-\rho^2)\}^{-2}$$

$$= 64(1+\rho^2) \; \{(1-\rho) \quad (1-\rho^2)\}^{-2} < \infty.$$

(iv) $\mathrm{Var}(\xi_n) = 1$ for all n,

and

(v) if F_{ni} is the distribution function of $Y_{i,n}$, then for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \sum_{i=1}^{n} \int_{|x|>\varepsilon} x^2 \, dF_{ni}(x) \le \lim_{n\to\infty} \varepsilon^{-1} \sum_{i=1}^{n} \int_{|x|>\varepsilon} |x|^3 \, dF_{ni}(x)$$

$$\le \varepsilon^{-1} \lim_{n\to\infty} \sum_{i=1}^{n} E|Y_{i,n}|^3$$

$$\le \varepsilon^{-1} C^3 \lim_{n\to\infty} s_n^{-3} n^{-3/2} \sum_{i=1}^{n} |\gamma_n^{i\cdot}|^3 E|X_i - C\mu_o|^3$$

$$= \varepsilon^{-1} C^3 (2\sqrt{2/\pi}) \lim_{n\to\infty} n^{-3/2} \sum_{i=1}^{n} |\gamma_n^{i\cdot}|^3$$

since $E|X_i - C\mu_o|^3 = 2\sqrt{2/\pi}$ and $\lim_{n\to\infty} s_n^2 = 1$. Hence, by (2.5.16),

$$\lim_{n\to\infty} \sum_{i=1}^{n} \int_{|x|>\varepsilon} x^2 \, dF_{ni}(x)$$

$$\le a \, \varepsilon^{-1} (1-\rho)^3 (2\sqrt{2/\pi}) \lim_{n\to\infty} n^{-1/2} \tag{2.5.26}$$

where $a = 512 \{(1-\rho)^6 (1-\rho^2)^3\}^{-1}$. Since the right side of (2.5.26) tends to zero as $n \to \infty$,

$$\lim_{n\to\infty} \sum_{i=1}^{n} \int_{|x|>\varepsilon} x^2 \, dF_{ni}(x) = 0 \tag{2.5.27}$$

for every $\varepsilon > 0$.

Thus all the conditions of Lemma 2.2.6 are satisfied. Hence ξ_n converges in distribution to normal with mean zero and variance one. Since s_n^2 tends to 2 as $n \to \infty$, we get that $n^{-1/2} \frac{\partial}{\partial\mu} \log p_n(\mu_o)$ converges to normal with mean zero and variance two. Therefore, by (2.5.24), since $\{n/I_n(\mu_o)\}^{1/2}$ tends to 1, we get that

$$Y_n = \{I_n(\mu_o)\}^{-1/2} \frac{\partial}{\partial\mu} \log p_n(\mu_o)$$

converges to normal with mean zero and variance two. Since the asymptotic distribution of $\{I_n(\mu_o)\}^{1/2} (\hat{\mu}_n-\mu_o)$ is the same as that of Y_n, we get that $\{I_n(\mu_o)\}^{1/2}(\hat{\mu}_n-\mu_o)$ converges to normal with mean zero and variance one. This proves asymptotic normality of the MLE $\hat{\mu}_n$ of μ.

2.6 Conclusions.

In this chapter, we proved weak consistency and first-order efficiency of a MLE for m-dependent sequences of random variables, under some regularity conditions. From definition 2.2.1, it follows that, when m is equal to zero, m-dependence reduces to independence of the random variables. Then clearly (2.3.1) becomes

$$p(x_1,\ldots,x_n;\theta) = \{\prod_{i=1}^{n} p(x_i;\theta)\}\ \phi_k(x_1,\ldots,x_n;\theta) \qquad (2.6.1)$$

so that $\phi_k(x_1,\ldots,x_n;\theta) \equiv 1$, for all k. It can easily be seen that all the conditions of Section 2.3 will be satisfied and hence Theorems 2.4.1 and 2.4.2 are valid in this case. In fact, the strong consistency and asymptotic normality of MLE hold in this case.

CHAPTER III

MAXIMUM LIKELIHOOD ESTIMATION FOR STATIONARY ϕ-MIXING PROCESSES

3.1 Introduction.

In recent years, several authors investigated the properties of stochastic processes satisfying certain "mixing conditions" (for example, see Ibragimov [20], Serfling [46] and the references there). Roughly speaking, any mixing condition says that the dependence between the random variables is weaker the farther they are apart, or else the dependence between the end of the process and its beginning is weak.

In this chapter, we establish weak consistency, asymptotic normality and first-order efficiency of an approximate maximum likelihood estimator (AMLE) when the observations on which the AMLE is based are from a stationary stochastic process satisfying a ϕ-mixing condition. Definitions of AMLE and ϕ-mixing process are given in the following sections.

Under the set of conditions presented in this chapter, the proof of weak consistency and asymptotic normality of AMLE depends on weak law of large numbers and central limit theorem for double sequences of random variables respectively. Rosén [40] proved a central limit theorem for double sequences of random variables, under some general assumptions. We obtain a particular case of this central limit theorem for stationary ϕ-mixing random variables in Section 3.4. In Section 3.3, we use Khintchine's method to obtain weak law of large numbers for double sequences of random variables. The proofs of weak consistency, asymptotic normality and first-order efficiency of an AMLE are given in Section 3.6. Section 3.7 includes an example which satisfies the regularity conditions given in Section 3.5.

3.2 Preliminary Results:

In this section, we give some definitions and lemmas which are needed to prove the main results of the chapter.

Let $\{X_n, n \geq 1\}$ be a sequence of random variables defined on a probability space. Let $\mathfrak{M}_a^b$ denote the σ-field generated by events of the form $\{(X_{i_1},\ldots,X_{i_k}) \in E\}$, where $a-1 < i_1 < \ldots < i_k < b+1$ and E is a k-dimensional Borel set.

Definition 3.2.1: A sequence $\{X_n, n \geq 1\}$ of random variables is said to be ϕ-mixing if for each $t(1 \leq t \leq \infty)$ and for each $n(n \geq 1)$,

$$|P(B|\mathfrak{M}_1^t) - P(B)| \leq \phi_n \tag{3.2.1}$$

with probability one for all events $B \in \mathfrak{M}_{t+n}^{\infty}$, where ϕ_n is a sequence of positive numbers tending to zero as $n \to \infty$.

Remark 3.2.1: Condition (3.2.1) is equivalent to

$$|P(A \cap B) - P(A)\, P(B)| \leq \phi_n\, P(A) \tag{3.2.2}$$

for every events $A \varepsilon \mathfrak{M}_1^t$ and $B \varepsilon \mathfrak{M}_{t+n}^{\infty}$. Proof of this result can be found in Ibragimov [20].

Remark 3.2.2: If $\{X_n, n \geq 1\}$ is a m-dependent sequence of random variables (see Definition 2.2.1), then (3.2.1) will be satisfied for some sequence $\{\phi_n\}$ of positive numbers such that $\phi_n = 0$ for all $n > m$. Thus m-dependence is a particular case of ϕ-mixing and hence the results of this chapter will be true for any m-dependent process satisfying the assumptions given in the next section.

Now, we state and prove some lemmas. Proofs of these lemmas can be found in the references given below.

Lemma 3.2.1: (Ibragimov [20]). Let $\{X_j, j \geq 1\}$ be a stationary ϕ-mixing process. Let the random variable f be measurable with respect to $\mathfrak{M}_1^k$ and f_j be obtained by setting $f_j \equiv f(x_j, \ldots, x_{j+k-1})$. Then the process $\{f_j, j \geq 1\}$ is also stationary and ϕ_f-mixing, where

$$\phi_{f_j, n} \leq \begin{cases} 1 & \text{for } n \leq k \\ \phi_{(n-k)} & \text{for } n > k \end{cases} \tag{3.2.3}$$

Lemma 3.2.2: (Serfling [46]). Let $\{X_i\}$ satisfy condition (3.2.1). Let ξ be any random variable measurable with respect to $\mathfrak{M}^{\infty}_{a+k}$ such that $E|\xi|^p < \infty$ for some $p > 1$. If $1 \leq \alpha \leq p$, then

$$E|E(\xi|\mathfrak{M}^{a}_{1}) - E(\xi)|^{\alpha} \\ \leq 2^{\alpha}\phi_k^{\alpha/q} \{E|\xi|^p\}^{\alpha/p}, \tag{3.2.4}$$

where $p^{-1} + q^{-1} = 1$.

Lemma 3.2.3: (Ibragimov [20], Billingsley [10]). For a sequence of random variables, if ξ is measurable $\mathfrak{M}^{k}_{1}$ and η is measurable $\mathfrak{M}^{\infty}_{k+n}$ $(n \geq 0)$, then $E|\xi|^p < \infty$, $E|\eta|^q < \infty$, and $p^{-1} + q^{-1} = 1$ imply

$$|E(\xi\eta) - E(\xi)\, E(\eta)| \leq 2\, \phi_n^{1/p} \{E|\xi|^p\}^{1/p} \{E|\eta|^q\}^{1/q}. \quad \ldots \tag{3.2.5}$$

Lemma 3.2.4: (Cramér [14]). Let ξ_1, ξ_2, ... be a sequence of random variables, with the distribution functions (d.f.) $F_1, F_2, \ldots$. Suppose that $F_n(x)$ tends to a d.f. $F(x)$ as $n \to \infty$. Let η_1, η_2, ... be another sequence of random variables, and suppose that η_n converges in probability to a constant C. Let

$$X_n = \xi_n + \eta_n; \; Y_n = \xi_n \eta_n \text{ and } Z_n = \xi_n/\eta_n.$$

Then the d.f. of X_n tends to $F(x-C)$. Further if $C > 0$, the d.f. of Y_n tends to $F(x/C)$ while that of Z_n tends to $F(Cx)$.

Lemma 3.2.5: For any random variables X , Y and Z,

$$E|E(Y|X) - E(Y)| \le E|E(Z|X) - E(Z)| + 2E|Y-Z| .$$

Proof: We have

$$E|E(Y|X) - E(Y)| \le E|E(Y-Z| X) - E(Y-Z)| + E|E(Z|X) - E(Z)|$$

$$\le 2E|Y-Z| + E|E(Z|X) - E(Z)| .$$

Before we proceed further, we shall say a few words about the Lévy metric. Full treatment of these concepts can be found in Gnendenko and Kolmogorov [17].

Let D be the metric space of all probability distributions on the real line, with the Lévy distance L(.,.) , that is,

$$L(F,G) = \inf(h|F(x-h) - h \le G(x) \le F(x+h)+h , \text{ all } x) . \qquad \ldots(3.2.6)$$

Then D is a complete space and convergence in the sense of the metric is equivalent to convergence in distribution. The following lemma is probably not new.

Lemma 3.2.6: Let X and Y be any two random variables. Then

$$L(\mathcal{L}(X+Y), \mathcal{L}(X)) \le E\{|Y|^{1+\delta}\}^{1/(2+\delta)} \qquad (3.2.7)$$

for any $\delta \ge 0$ where $\mathcal{L}(X)$ denotes the distribution of X.

Proof: We have

$$F_{X+Y}(x) = P(X+Y\leq x)$$

$$\leq P(|Y| \leq h\,,\ X\leq x+h) + P(|Y|>h)$$

$$\leq F_X(x+h) + P(|Y|>h)\,. \qquad (3.2.8)$$

By substituting X+Y, -Y and x-h in the place of X, Y and x in (3.2.8) respectively, we get

$$F_X(x-h) \leq F_{X+Y}(x) + P(|Y|>h)\,. \qquad (3.2.9)$$

From (3.2.8) and (3.2.9) we get

$$F_X(x-h) - P(|Y|>h) \leq F_{X+Y}(x) \leq F_X(x+h) + P(|Y|>h) \qquad \ldots(3.2.10)$$

for all x. Hence, from (3.2.6) it follows that

$$L(\mathcal{L}(X+Y), \mathcal{L}(X)) \leq h \qquad (3.2.11)$$

if h satisfies $h \geq P(|Y|>h)$. An application of Chebyshev's inequality yields that this is fulfilled if

$$h \leq \{E|Y|^{1+\delta}\}^{1/(2+\delta)}$$

for any $\delta > 0$. Hence the lemma follows from (3.2.11).

The following three lemmas are extensions of Theorem 2.3, Lemma 2.1 and Theorem 3.1 of Serfling [46] to double sequences of random variables. The proofs are similar to the respective results in Serfling [46].

Lemma 3.2.7: Let $\{X_{n,i},\ i = 1,\ldots,\ \beta_n,\ n \geq 1\}$ be a double sequence of random variables satisfying the following conditions.

(i) $\beta_n \to \infty$, as $n \to \infty$.

(ii) The random variables in the nth row are stationary and $\phi^{(n)}$-mixing, where

$$\phi_k^{(n)} \leq \psi_k \tag{3.2.12}$$

for all $n \geq 1$; $1 \leq k \leq \beta_n$.

(iii) $E(X_{n,i}) = 0$ for all n and i.

(iv) $\overline{\lim_{n\to\infty}}\, E|X_{n,1}|^{2+\delta} = M < \infty$ for some δ such that $0 < \delta \leq 1$.

(v) $\psi_n = O(n^{-\theta})$, for some $\theta > 1 + \frac{2}{\delta}$ where δ is as in (iv).

Then, there exists positive constants C, ℓ and ε such that

$$E|E\{T^2_{\alpha\beta_n,\Delta} | S_{\alpha,n}\} - E(T^2_{\alpha\beta_n,\Delta})|$$

$$\leq C(\Delta^2 + \Delta^{1/2})\beta_n^{-\ell}\{E|X_{n,1}|^{2+\varepsilon}\}^{2/(2+\varepsilon)}$$

for all n, where $T_{a,\Delta} = \beta_n^{-1/2} \sum_{i=a+1}^{a+\Delta\beta_n} X_{n,i}$, $S_{\alpha,n} = T_{o,\alpha}$ and $0 \leq \alpha < \alpha+\Delta \leq 1$.

Proof: By assumption (ii) and Minkowski inequality, we have, for all $\varepsilon > 0$,

$$E|S_{\Delta,n}|^{2+\varepsilon} \leq \Delta^{2+\varepsilon}\, \beta_n^{1+(\varepsilon/2)}\, E|X_{n,1}|^{2+\varepsilon}\ . \tag{3.2.13}$$

Let $\{\gamma_n\}$ be a sequence of positive integers such that $\gamma_n = \beta_n^{\lambda}$, $0 < \lambda < 1$.

$$
\begin{aligned}
& E|E(T^2_{\alpha\beta_n,\Delta}|S_{\alpha,n}) - E(T^2_{\alpha\beta_n,\Delta})| \\
& \le E|E(T^2_{\alpha\beta_n+\gamma_n,\Delta}|S_{\alpha,n}) - E(T^2_{\alpha\beta_n+\gamma_n,\Delta})| \\
& \quad +2E|T^2_{\alpha\beta_n,\Delta} - T^2_{\alpha\beta_n+\gamma_n,\Delta}|
\end{aligned}
\tag{3.2.14}
$$

by Lemma 3.2.5. By Lemma 3.2.2, assumptions (i) and (ii) the right side of (3.2.14) is

$$
\begin{aligned}
& \le 2\{\phi^{(n)}_{\gamma_n}\}^{\varepsilon/(2+\varepsilon)} \{E|T_{\alpha\beta_n+\gamma_n,\Delta}|^{2+\varepsilon}\}^{2/(2+\varepsilon)} \\
& \quad + 2E\{|T_{\alpha\beta_n,\Delta} + T_{\alpha\beta_n+\gamma_n,\Delta}| \, |T_{\alpha\beta_n,\Delta} - T_{\alpha\beta_n+\gamma_n,\Delta}|\} \\
& \le 2\{\psi_{\gamma_n}\}^{\varepsilon/2+\varepsilon} \{E|S_{\Delta,n}|^{2+\varepsilon}\}^{2/(2+\varepsilon)} \\
& \quad + 2E\{|T_{\alpha\beta_n,\Delta} + T_{\alpha\beta_n+\gamma_n,\Delta}| \, |R_{\alpha,n} - R_{\alpha+\Delta,n}|\}
\end{aligned}
\tag{3.2.15}
$$

where $R_{\alpha,n} = \beta_n^{-1/2} \sum_{i=\alpha\beta_n+1}^{\alpha\beta_n+\gamma_n} X_{n,i}$. The right side of (3.2.15) is

$$
\begin{aligned}
& \le 2\{\psi_{\gamma_n}\}^{\varepsilon/(2+\varepsilon)} \{E|S_{\Delta,n}|^{2+\varepsilon}\}^{2/(2+\varepsilon)} \\
& \quad + 2[\{E|T_{\alpha\beta_n,\Delta} + T_{\alpha\beta_n+\gamma_n,\Delta}|^2\}\{E|R_{\alpha,n} - R_{\alpha+\Delta,n}|^2\}]^{1/2}
\end{aligned}
$$

$$\le 2\{\psi_{\gamma_n}\}^{\varepsilon/(2+\varepsilon)} \{E|S_{\Delta,n}|^{2+\varepsilon}\}^{2/(2+\varepsilon)}$$

$$+ 8[\{E|T_{\alpha\beta_n,\Delta}|^2\} \{E|R_{\alpha,n}|^2\}]^{1/2}$$

by stationarity and C_r-inequality (Loeve [28],

Therefore,

$$|E(T^2_{\alpha\beta_n,\Delta}|\ S_{\alpha,n}) - E(T^2_{\alpha\beta_n,\Delta})|$$

$$\le 2\{\psi_{\gamma_n}\}^{\varepsilon/(2+\varepsilon)}\{ E|S_{\Delta,n}|^{2+\varepsilon}\}^{2/(2+\varepsilon)}$$

$$+ 8[\{E|S_{\Delta,n}|^2\}\{E(\beta_n^{-1/2} \sum_{i=1}^{\gamma_n} X_{n,i})^2\ ]^{1/2} .$$

Now consider

$$E[\beta_n^{-1/2} \sum_{i=1}^{\Delta\beta_n} X_{n,i}]^2 = \beta_n^{-1}[\sum_{i=1}^{\Delta\beta_n} E(X^2_{n,i}) + 2 \sum_{i=1}^{\Delta\beta_n} \sum_{j=i+1}^{\Delta\beta_n} E(X_{n,i}X_{n,j})]$$

$$\le \beta_n^{-1}[\Delta\beta_n\ E(X^2_{n,1}) + 2 \sum_{i=1}^{\Delta\beta_n} \sum_{j=i+1}^{\Delta\beta_n} |E(X_{n,1}\ X_{n,j-i+1})|]$$

$$\le \Delta E(X^2_{n,1}) + 2\beta_n^{-1} \sum_{i=1}^{\Delta\beta_n} \sum_{j=i+1}^{\Delta\beta_n} \{\ \phi^{(n)}_{j-i}\}^{1/2}\ E(X^2_{n,1}) \quad (3.2.17)$$

by Lemma 3.2.3 and assumptions (ii) and (iii). Therefore, by (3.2.12) we get

$$E[\beta_n^{-1/2} \sum_{i=1}^{\Delta\beta_n} X_{n,i}]^2 \le \Delta E(X^2_{n,1}) + 2\beta_n^{-1} \sum_{i=1}^{\Delta\beta_n} \sum_{k=1}^{\infty} \psi_k^{1/2}\ E(X^2_{n,1})$$

$$= \Delta E(X^2_{n,1}) + 2\beta_n^{-1}\ E(X^2_{n,1}) \sum_{i=1}^{\Delta\beta_n} \sum_{k=1}^{\infty} \{O(k^{-\theta})\}^{1/2} \quad (3.2.18)$$

by assumption (v). That is,

$$E(\beta_n^{-1/2} \sum_{i=1}^{\Delta\beta_n} X_{n,i})^2 \leq \Delta E(X_{n,1}^2) + 2\Delta M_2 \, E(X_{n,1}^2) \qquad (3.2.19)$$

where

$$M_2 = \sum_{k=1}^{\infty} \{O(k^{-\theta})\}^{1/2} < \infty. \qquad (3.2.20)$$

Note that $\theta > 2$ since $\delta < .1$. Now, by (3.2.13), (3.2.16) and (3.2.19) we get

$$|E(T^2_{\alpha\beta_n,\Delta}|S_{\alpha,n}) - E(T^2_{\alpha\beta_n,\Delta})|$$

$$\leq 2\{\psi_{\gamma_n}\}^{\varepsilon/(2+\varepsilon)}\{\Delta^{2+\varepsilon}\beta_n^{1+\varepsilon/2} \, E|X_{n,1}|^{2+\varepsilon}\}^{2/(2+\varepsilon)}$$

$$+ 8[\{(1+2M_2)\Delta \, E(X_{n,1}^2)\}\{(1+2M_2)\,\gamma_n\beta_n^{-1} \, E(X_{n,1}^2)\}]^{1/2}.$$

Then, since $\gamma_n = \beta_n^{\lambda}$ and $0 < \lambda < 1$, by assumption (v), the right side is

$$\leq 2M_3 \, \beta_n^{-\lambda\theta\varepsilon/(2+\varepsilon)} \, \Delta^2 \, \beta_n^{2\gamma/(2+\varepsilon)} \, \{E|X_{n,1}|^{2+\varepsilon}\}^{2/(2+\varepsilon)}$$

$$+ 8(1+2M_2) \, \Delta^{1/2} \, \beta_n^{-(1-\lambda)/2} \, E(X_{n,1}^2)$$

where M_3 is a constant independent of n and $2\gamma = 2+\varepsilon$.

Therefore, we have, by Holder's inequality,

$$|E(T^2_{\alpha\beta_n,\Delta}|S_{\alpha,n}) - E(T^2_{\alpha\beta_n,\Delta})|$$

$$\leq 2M_3\beta_n^{-(\lambda\theta\varepsilon-2\gamma)/2\gamma}\Delta^2 \, \{E|X_{n,1}|^{2\gamma}\}^{1/\gamma}$$

$$+ 8(1+2M_2) \, \Delta^{1/2} \, \beta_n^{-(1-\lambda)/2} \, \{E|X_{n,1}|^{2\gamma}\}^{1/\gamma}. \qquad (3.2.21)$$

Since $\theta > 1 + 2/\delta$, there exists an ε such that $0 < \varepsilon \leq \delta$ and $\theta > 1 + 2/\varepsilon$, that is, $\theta\varepsilon > 2\gamma$. For this ε, if we choose $\lambda = 3\gamma/(\theta\varepsilon+\gamma)$, then $0 < \lambda < 1$ and (3.2.21) will imply that

$$|E(T^2_{\alpha\beta_n,\Delta}|S_{\alpha,n}) - E(T^2_{\alpha\beta_n,\Delta})|$$

$$\leq 2M_3\ \beta_n^{-(\theta\varepsilon-2\gamma)/(2\theta\varepsilon+2\gamma)}\ \Delta^2\ \{E|X_{n,1}|^{2\gamma}\}^{1/\gamma}$$

$$+ 8(1+2M_2)\ \Delta^{1/2}\ \beta_n^{-(\theta\varepsilon-2\gamma)/(2\theta\varepsilon+2\gamma)}\ \{E|X_{n,1}|^{2\gamma}\}^{1/\gamma}$$

$$= C\beta_n^{-\ell}\ (\Delta^2+\Delta^{1/2})\ \{E|X_{n,1}|^{2+\varepsilon}\}^{2/(2+\varepsilon)} \qquad (3.2.22)$$

where

$$C = 2M_3 + 8(1+2M_2)\ , \qquad (3.2.23)$$

$$\ell = (\theta\varepsilon-2\gamma)/(2\theta\varepsilon+2\gamma), \qquad (3.2.24)$$

$$\gamma = 1 + \varepsilon/2\ , \qquad (3.2.25)$$

and

$$0 < \varepsilon \leq \delta. \qquad (3.2.26)$$

Hence from (3.2.22) - (3.2.26) the lemma follows.

Lemma 3.2.8: Under the assumptions of the previous lemma, there exist positive constants C'_β and β such that for every n,

$$E|E(T^2_{\alpha\beta_n,\Delta}\ |S_{\alpha,n}) - E(T^2_{\alpha\beta_n,\Delta})|^{1+\beta} \leq C'_\beta\ .$$

26097

Proof: Let $Y_{\alpha\beta_n,\Delta} = E(T^2_{\alpha\beta_n\Delta} | S_{\alpha,n}) - E(T^2_{\alpha\beta_n,\Delta})$. By C_r-inequality, assumption (ii) implies that

$$E|Y_{\alpha\beta_n,\Delta}|^{1+\delta/2} \leq 2^{1+\delta/2} E|T_{\alpha\beta_n,\Delta}|^{2+\delta}$$

$$= 2^{1+\delta/2} E|S_{\Delta,n}|^{2+\delta}$$

$$\leq 2^{1+\delta/2} \Delta^{2+\delta} \beta_n^{1+\delta/2} E|X_{n,1}|^{2+\delta}. \quad (3.2.27)$$

The last step follows by Minkowski inequality. Let β be such that $0 < \beta < \ell\delta/(2+2\ell+\delta)$, where ℓ is as in (3.2.24). Let A_n denote the event $\{|Y_{\alpha\beta_n,\Delta}|^{\beta} \leq r_n^{-1}\}$ and A_n' its complement, where

$$r_n = C\, \beta_n^{-\ell}\{E|X_{n,1}|^{2+\varepsilon}\}^{2/(2+\varepsilon)} (\Delta^2+\Delta^{1/2}) \quad (3.2.28)$$

Taking the expectation of $|Y_{\alpha\beta_n,\Delta}|^{1+\beta}$ over A_n and A_n' separately, it follows that

$$E|Y_{\alpha\beta_n,\Delta}|^{1+\beta} \leq r_n^{-1} E|Y_{\alpha\beta_n,\Delta}| + r_n^{(\delta-2\beta)/2\beta} E|Y_{\alpha\beta_n,\Delta}|^{1+\delta/2}. \quad \ldots(3.2.29)$$

By (3.2.22), we have

$$r_n^{-1} E|Y_{\alpha\beta_n,\Delta}| \leq 1\,. \quad (3.2.30)$$

The second term on the right side of (3.2.29) is

$$\leq r_n^{(\delta-2\beta)/2\beta}\, 2^{1+\delta/2} \Delta^{2+\delta} \beta_n^{1+\delta/2} E|X_{n,1}|^{2+\delta} \quad (3.2.31)$$

by (3.2.27). Since $(\delta-2\beta)/(2\beta) > (2+\delta)/(2\ell)$, $0 < \varepsilon \leq \delta$ and $0 < \Delta \leq 1$, by (3.2.28), we get

$$r_n^{(\delta-2\beta)/2\beta} < C_\beta\, \beta_n^{-(1+\delta/2)} \{E|X_{n,1}|^{2+\delta}\}^{2(\delta-2\beta)/(2+\delta)2\beta} \qquad \ldots(3.2.32)$$

where C_β is a positive constant which depends on β. Therefore, from (3.2.29)-(3.2.32), we get

$$\begin{aligned} E|Y_{\alpha\beta_n,\Delta}|^{1+\beta} &\leq 1 + C_\beta\, 2^{1+\delta/2}\, \Delta^{2+\delta}\{E|X_{n,1}|^{2+\delta}\}^{(1+\beta)/\beta} \\ &\leq 1 + C_\beta\, 2^{1+\delta/2}\, M^{(1+\beta)/\beta} \qquad (3.2.33) \end{aligned}$$

for $n \geq N_o$, by assumption (iv) of Lemma 3.2.7. Let

$$d = \max\{E|X_{1,1}|^{2+\delta}, E|X_{2,1}|^{2+\delta}, \ldots, E|X_{N_o,1}|^{2+\delta}, M\}.$$

Then,

$$\begin{aligned} E|Y_{\alpha\beta_n,\Delta}|^{1+\beta} &\leq 1 + C_\beta\, 2^{1+\delta/2}\, d^{(1+\beta)/\beta} \\ &= C'_\beta \qquad (3.2.34) \end{aligned}$$

for all n, where

$$C'_\beta = 1 + C_\beta\, 2^{1+\delta/2}\, d^{(1+\beta)/\beta}. \qquad (3.2.35)$$

Hence the lemma follows.

Remark 3.2.1: Note that the estimate in the lemma holds for all β such that $0 < \beta < \ell\delta/(2+2\ell+\delta)$.

Lemma 3.2.9: If the assumptions of Lemma 3.2.7 are satisfied then there exists a τ such that $0 < \tau \leq \delta$ and L_τ such that

$$E\left|\sum_{i=1}^{\beta_n} X_{n,i}\right|^{2+\tau} \leq \beta_n^{1+\tau/2} L_\tau$$

for every $n > N_\tau$.

Proof: Since $2\ell\delta/(2+2\ell+\delta) > 0$ in Lemma 3.2.8, there exists a positive number τ such that

$$\tau/2 < \ell\cdot\delta\ (2+2\ell+\delta)^{-1} . \tag{3.2.36}$$

We take β in Lemma 3.2.8 to be $\tau/2$. Without loss of generality we can assume that

$$\beta \leq \min[\frac{1}{2}, \frac{\delta}{2}]$$

so that

$$0 < \tau \leq \min[1, \delta] . \tag{3.2.37}$$

Let $m = [\frac{\beta n}{2}]$, the greatest integer $\leq \beta_n/2$ and define

$$R_{n,m} = \sum_{i=1}^{m} X_{n,i} \quad \text{and} \quad S_{n,m} = \sum_{i=m+1}^{2m} X_{n,i} .$$

Since $\tau \leq 1$, we have, by C_r-inequality, that

$$\begin{aligned} E|R_{n,m} + S_{n,m}|^{2+\tau} &\leq E\{(R_{n,m}+S_{n,m})^2(|R_{n,m}|^{\tau}+|S_{n,m}|^{\tau})\} \\ &\leq E|R_{n,m}|^{2+\tau} + E|S_{n,m}|^{2+\tau} + 2E\{|R_{n,m}||S_{n,m}|^{1+\tau}\} \\ &\quad + 2E\{|R_{n,m}|^{1+\tau}|S_{n,m}|\} + E\{|R_{n,m}|^2|S_{n,m}|^{\tau}\} \\ &\quad + E\{|R_{n,m}|^{\tau}\ |S_{n,m}|^2\} . \end{aligned} \tag{3.2.38}$$

Now, letting $\Delta_{n,m} = E(S_{n,m}^2|\ R_{n,m}) - E(S_{n,m}^2)$, we have for $0 < s \leq 2$,

$$E(|S_{n,m}|^s\ |R_{n,m}) \leq \{E|S_{n,m}|^2|R_{n,m}\}^{s/2}$$

by Holder's inequality. Then

$$E(|S_{n,m}|^s|R_{n,m}) \leq \{\Delta_{n,m} + E(S_{n,m})^2\}^{s/2}$$

$$\leq |\Delta_{n,m}|^{s/2} + \{E(S_{n,m})^2\}^{s/2}$$

by C_r-inequality. Hence, for $r+s = 2+\tau$ and $0 < s \leq 2$,

$$E|R_{n,m}|^r\ |S_{n,m}|^s \leq E\{|R_{n,m}|^r\ E(|S_{n,m}|^s\ |R_{n,m})\}$$

$$\leq \{E|R_{n,m}|^{2+\tau}\}^{r/(2+\tau)}\ [\{E(S^2_{n,m})\}^{s/2}$$

$$+ \{E|\ \Delta_{n,m}|^{1+\tau/2}\}^{s/(2+\tau)}]$$

by Holder's inequality. Therefore,

$$E|R_{n,m}|^r|S_{n,m}|^s \leq \{E|R_{n,m}|^{2+\tau}\}^{r/(2+\tau)} [\{(2M_2+1)\ m\ E(X^2_n$$

$$+ \{C'_{\tau/2}\}^{s/(2+\tau)}]$$

by stationarity, (3.2.19) with $\Delta = m\ \beta_n^{-1}$ and Lemma 3.2.8. By assumption (iv) of Lemma 3.2.7 for sufficiently large n, say $n > N_o$, we get

$$E|R_{n,m}|^r|S_{n,m}|^s \leq \{E|R_{n,m}|^{2+\tau}\}^{r/(2+\tau)}\ [C_1\ m^{s/2} + C_{2,\tau}]$$

where $C_{2,\tau} = \{C'_{\tau/2}\}^{s/(2+\tau)}$. Therefore, for $n > N_1$, $(N_1 \geq N_o)$

$$E|R_{n,m}|^r|S_{n,m}|^s \leq C_2\{E|R_{n,m}|^{2+\tau}\}^{r/(2+\tau)}\ m^{s/2} \qquad (3.2.40)$$

where C_2 depends on s. Define for positive integers h,

$$A_{h,n} = h^{-(1+\tau/2)}\ E|\sum_{i=1}^{h} X_{n,i}|^{2+\tau} \qquad (3.2.41)$$

which is finite by assumption (iv) of Lemma 3.2.7 since $\tau \leq \delta$. Then from (3.2.38), there exists a constant C_2' such that for large n (say) $n > N'(N' \geq N_1)$

$$
\begin{aligned}
E|R_{n,m} + S_{n,m}|^{2+\tau} \leq\ & 2E|R_{n,m}|^{2+\tau} \\
& + C_2'[2\{E|R_{n,m}|^{2+\tau}\}^{1/(2+\tau)}\ m^{(1+\tau)/2} \\
& \quad + 2\{E|R_{n,m}|^{2+\tau}\}^{(1+\tau)/(2+\tau)} m^{1/2} \\
& \quad + \{E|R_{n,m}|^{2+\tau}\}^{2/(2+\tau)}\ m^{\tau/2} \\
& \quad + \{E|R_{n,m}|^{2+\tau}\}^{\tau/(2+\tau)}\ m] \\
=\ & 2m^{1+\tau/2}\ A_{m,n} \\
& + C_2'[2\{A_{m,n}\ m^{1+\tau/2}\}^{1/(2+\tau)}\ m^{(1+\tau)/2} \\
& + 2\{A_{m,n}\ m^{1+\tau/2}\}^{(1+\tau)/(2+\tau)}\ m^{1/2} \\
& \quad + \{A_{m,n}\ m^{1+\tau/2}\}^{2/(2+\tau)}\ m^{\tau/2} \\
& \quad + \{A_{m,n}\ m^{1+\tau/2}\}^{\tau/(2+\tau)}\ m] \\
\leq\ & m^{1+\tau/2}\ A_{m,n}[2+g(A_{m,n})]
\end{aligned}
\tag{3.2.42}
$$

where

$$g(Z) = C_2'[2Z^{-(1+\tau)/(2+\tau)} + 2Z^{-1/(2+\tau)} + Z^{-\tau/(2+\tau)} + Z^{-2/(2+\tau)}].$$

Since $g(Z) \to 0$, as $Z \to \infty$, there exists $Z_o > 1$ such that

$$\{2+g(Z)\}^{1/(2+\tau)} \leq 2^{1/2} - Z_o^{-1} \tag{3.2.43}$$

for $Z \geq Z_o$. Define

$$a_{h,n} = \max(A_{h,n}, Z_o, M) \tag{3.2.45}$$

where M is as defined in assumption (iv) of Lemma 3.2.7. Since $Z(2+g(Z))$ is a non-decreasing function of Z, we have

$$A_{m,n}[2+g(A_{m,n})] \leq a_{m,n}[2+g(a_{m,n})].$$

Hence from (3.2.42) and (3.2.43) it follows that

$$E|R_{n,m} + S_{n,m}|^{2+\tau} \leq m^{1+\tau/2} \; a_{m,n}((2^{1/2}-Z_o^{-1})^{2+\tau} .$$

Since $\beta_n - 1 \leq 2m \leq \beta_n$, we obtain, by Minkowski inequality that

$$\begin{aligned}[E|\sum_{i=1}^{\beta_n} X_{n,i}|^{2+\tau}]^{1/(2+\tau)} &\leq \{E|R_{n,m} + S_{n,m}|^{2+\tau}\}^{1/(2+\tau)} \\ &\quad + \{E|X_{\beta_n, n}|^{2+\tau}\}^{1/(2+\tau)} . \\ &\leq \{E|R_{n,m} + S_{n,m}|^{2+\tau}\}^{1/(2+\tau)} + M \\ &\leq m^{1/2} \; a_{m,n}^{1/(2+\tau)} \; (2^{1/2} - Z_o^{-1}) + a_{m,n}^{1/(2+\tau)} \\ &\leq m^{1/2} \; a_{m,n}^{1/(2+\tau)} \; (2^{1/2} - Z_o^{-1} + m^{-1/2}).\end{aligned}$$

Choose n sufficiently large, say $n > N_2$ so that $m^{-1/2} < Z_o^{-1}$. Then for $n > \max(N', N_2)$

$$\begin{aligned}E|\sum_{i=1}^{\beta_n} X_{n,i}|^{2+\tau} &\leq (2m)^{1+\tau/2} \; a_{m,n} \\ &\leq \beta_n^{1+\tau/2} \; a_{m,n} \quad .\end{aligned} \tag{3.2.45}$$

It follows from (3.2.45) and the definition of $A_{\beta_n,n}$ that $a_{\beta_n,n} \leq a_{m,n}$, for all $n > \max(N', N_2)$. Let $N = \beta_{\max(N',N_2)}$.

Then $a_{\beta_n,n} \leq a_{m,n}$ for all n such that $\beta_n > N$. Suppose n is such that $N < \beta_n \leq 2N$. Then $\frac{N}{2} < m \leq N$ and hence

$$a_{\beta_n,n} \leq a_{m,n} \leq \max(a_{1,n},\ldots, a_{N,n}). \qquad (3.2.46)$$

Now, we suppose that if n is such that $2^{k-1}N < \beta_n \leq 2^k N$, then

$$a_{\beta_n,n} \leq \max(a_{1,n},\ldots,a_{N,n}) \qquad (3.2.47)$$

and consider all those n's for which $2^k N < \beta_n \leq 2^{k+1} N$. Then $2^{k-1} N < m \leq 2^k N$ so that, from (3.2.47) it follows that

$$a_{m,n} \leq \max(a_{1,n},\ldots, a_{N,n}).$$

Hence by (3.2.46), we get

$$a_{\beta_n,n} \leq \max(a_{1,n},\ldots, a_{N,n})$$

for all n satisfying $\beta_n > N$. Therefore, for $\beta_n > N$

$$E|\sum_{i=1}^{\beta_n} X_{n,i}|^{2+\tau} \leq \beta_n^{1+\tau/2} \max(a_{1,n},\ldots,a_{N,n})$$

$$= \beta_n^{1+\tau/2} a_{k,n} \text{ (say)}$$

for some k such that $1 \leq k \leq N$. Since $\tau \leq \delta$, for all n such that $\beta_n > N$

$$A_{k,n} = k^{-(1+\tau/2)} E|\sum_{i=1}^{k} X_{n,i}|^{2+\tau}$$

$$\leq k^{1+\tau/2} E|X_{n,1}|^{2+\tau}$$

by Minkowskii inequality and stationarity. Hence

$$A_{k,n} \leq k^{1+\tau/2} M$$

by assumption (iv) of Lemma 3.2.7. Let

$$L = \max (MN^{1+\tau/2}, Z_o, M) .$$

Then L is finite for all n such that $\beta_n > N$ and

$$a_{k,n} = \max(A_{k,n}, Z_o, M) \leq L_\tau .$$

Hence

$$E\left| \sum_{i=1}^{\beta_n} X_{n,i} \right|^{2+\tau} \leq \beta_n^{1+\tau/2} L_\tau \tag{3.2.48}$$

for all n such that $\beta_n > N$. We note here that this N may depend on τ. From (3.2.48) it follows that

$$E\left| \sum_{i=1}^{\Delta\beta_n} X_{n,i} \right|^{2+\tau} \leq \Delta^{1+\tau} \beta_n^{1+\tau/2} L_\tau \tag{3.2.49}$$

for all β_n greater than some N. (This N need not be same as the above N). Thus the lemma follows.

Definition 3.2.2: A function g(x) is said to be uniformly integrable with respect to a family of distribution functions $\{F_n, n \geq 1\}$, if

$$\lim_{C\to\infty} \limsup_{n\to\infty} \int_{|x|>C} |g(x)| \, dF_n(x) = 0. \tag{3.2.50}$$

Definition 3.2.3: A family Π of distribution functions is said to be relatively compact if every sequence of elements of Π contains a convergent subsequence, converging to a distribution function.

Lemma 3.2.10: (Loéve [28]). If for a given $r_o > 0$, $|x|^{r_o}$ is uniformly integrable with respect to the family $\{F_n, n \geq 1\}$, then the sequence $\{F_n, n \geq 1\}$ is relatively compact.

Proof: For proof, we refer to Loéve [28], p.184.

3.3 Weak Law of Large Numbers for Double Sequences of Random Variables:

Let $\{Y_{n,i}, i = 1,2,\ldots,\beta_n, n \geq 1\}$ be a double sequenee sequence of random variables. We shall assume that $\beta_n \to \infty$ as $n \to \infty$. We first introduce some notation, related to the double sequence, which will be used throughout the later sections.

When we put a non-integer λ in a position where there should be an integer, then λ is interpreted as its integral part. Empty summation yields zero. $\mathcal{L}(X)$ stands for the distribution of the random variable X and is also used to denote the distribution function $F_X(x) = P(X \leq x)$. $E(X|\mathcal{B})$ stands for the conditional mean of X, given the σ-field $\mathcal{B}$. We usually write $E(X|Y)$ instead of $E(X|\mathcal{B}(Y))$, where $\mathcal{B}(Y)$ is the σ-field generated by the random variable Y. $N(\mu,\sigma^2)$ stands for the normal distribution with mean μ and variance σ^2. We write

$F_{n,i} = \mathcal{L}(Y_{n,i})$; $\sigma^2_{ni} = \sigma^2(Y_{n,i})$;

$$S_{\alpha,n} = Y_{n,1} + Y_{n,2} + \dots + Y_{n,\alpha\beta_n},\ 0 \le \alpha \le 1\ ;$$

$$T^n_{a,\Delta} = \sum_{i=a+1}^{a+\Delta\beta_n} Y_{n,i},\ 0 \le \Delta \le 1\ ;$$

and

$$G^n_{\alpha,\Delta} = \mathcal{L}(T^n_{\alpha\beta_n,\Delta}),\ 0 \le \alpha < \alpha+\Delta \le 1.$$

We denote $G^n_{0,\alpha}$ by $G_{\alpha,n}$ and $G_{0,n}$ by δ_0, where δ_μ stands for the distribution which has its entire mass at the point μ. In fact, $S_{\alpha,n} = T^n_{0,\alpha}$ and $G_{\alpha,n} = \mathcal{L}(S_{\alpha,n})$.

Now, we state and prove the weak law of large numbers for double sequences of random variables. The proof is similar to the proof of central limit theorem obtained by Rosén [40].

Theorem 3.3.1: Let $\{Y_{n,i}\ i = 1,2,\dots,\beta_n;\ n \ge 1\}$ be a double sequence of random variables which satisfies the following conditions:

(i) The random variables in the same row are defined on the same probability space.

(ii) $\overline{\lim\limits_{n\to\infty}}\, E|S_{\alpha,n} - S_{\gamma,n}|^{1+\delta} \le \chi(\alpha-\gamma),\ 0 \le \gamma < \alpha \le 1$ for some $\delta>0$ and for some function $\chi(s)$ which is bounded on $0 \le s \le 1$ and tends to zero as s tends to zero.

(iii) $\lim\limits_{\Delta\to+0} \Delta^{-1}\, \overline{\lim\limits_{n\to\infty}}\, E|E(T^n_{\alpha\beta_n,\Delta}|S_{\alpha,n})| = 0,\ 0 \le \alpha < 1.$

(iv) $\lim\limits_{\Delta\to+0} \Delta^{-1}\, \overline{\lim\limits_{n\to\infty}} \int_{|x|>\varepsilon} x\, dG^n_{\alpha,\Delta}(x) = 0,\ 0 \le \alpha < 1$ for every $\varepsilon > 0$.

Then for $0 \leq \alpha \leq 1$,

$$(p)\lim_{n\to\infty} S_{\alpha,n} = 0. \qquad (3.3.1)$$

We assume that

$$\lim_{n\to\infty} G_{\alpha,n} = G_{\alpha}, \qquad (3.3.2)$$

for $0 \leq \alpha \leq 1$, and afterwards show that this condition is superfluous. The proof of the theorem depends on the following lemmas. The proofs of all these lemmas and the theorem are similar to those of the corresponding results of Rosén [40].

Lemma 3.3.1: When condition (ii) is satisfied, the family $\{G_{\alpha,n}(x), n \geq 1\}$ integrates $|x|^{\varepsilon}$ uniformly for $0 \leq \varepsilon < 1+\delta$ and for fixed α.

Proof: By putting $\gamma = 0$ in condition (ii) we get that

$$\overline{\lim_{n\to\infty}} E|S_{\alpha,n}|^{1+\delta} < \infty .$$

Then

$$\int_{|x|>C} |x|^{\varepsilon} dG_{\alpha,n}(x) \leq C^{-(1+\delta-\varepsilon)} \int_{|x|>C} |x|^{1+\delta}\, dG_{\alpha,n}(x)$$

$$\leq C^{-(1+\delta-\varepsilon)}\, E|S_{\alpha,n}|^{1+\delta}. \qquad (3.3.3)$$

Therefore

$$\lim_{C\to\infty} \sup_{n} \int_{|x|>C} |x|^{\varepsilon}\, dG_{\alpha,n}(x) = 0$$

so that the lemma follows.

Lemma 3.3.2: Under assumptions (ii) and (3.3.2), the family $\{G_{\alpha}(x), 0 \leq \alpha \leq 1\}$ is continuous in the Lévy metric.

Proof: Let $\gamma < \alpha$. Then

$$L(G_\alpha, G_\gamma) = \lim_{n\to\infty} L(G_{\alpha,n}, G_{\gamma,n})$$

$$= \lim_{n\to\infty} L(\mathcal{L}(S_{\alpha,n}), \mathcal{L}(S_{\gamma,n}))$$

$$= \lim_{n\to\infty} L(\mathcal{L}(X_n+Y_n), \mathcal{L}(X_n))$$

where $X_n = S_{\gamma,n}$ and $Y_n = S_{\alpha,n} - S_{\gamma,n}$. Therefore by Lemma 3.2.6, we get

$$L(G_\alpha, G_\gamma) \leq \{\overline{\lim_{n\to\infty}}\, E|Y_n|^{1+\delta}\}^{1/(2+\delta)}$$

$$\leq \{\chi(\alpha-\gamma)\}^{1/(2+\delta)}$$

by condition (ii). Since $\chi(\alpha-\gamma) \to 0$ as $\alpha-\gamma \to 0$, the lemma follows.

Lemma 3.3.3: If assumptions (ii) and (3.3.2) are satisfied, then the function $\phi(t,\alpha)$, which is the characteristic function of G_α, satisfies

(a) $\phi(t,\alpha)$ is continuous for $0 \leq \alpha \leq 1$, $-\infty < t < \infty$; and

(b) $\frac{\partial}{\partial t}\phi(t,\alpha)$ exists and is continuous for $0 \leq \alpha \leq 1$, $-\infty < t < \infty$.

Proof: Since the characteristic functions converge uniformly on every compact t-interval whenever the corresponding distribution functions converge, (a) follows from Lemma 3.3.2.

According to assumption (ii) and Fatou's lemma, it holds that, for some $\delta > 0$,

$$\int |x|^{1+\delta} dG_\alpha(x) \leq \overline{\lim_{n\to\infty}} \int |x|^{1+\delta} dG_{\alpha,n}(x)$$

$$\leq \sup_{0\leq\alpha\leq 1} \chi(\alpha) < \infty.$$

Thus $\frac{\partial}{\partial t}\phi(t,\alpha)$ exists and

$$\frac{\partial}{\partial t}\phi(t,\alpha) = \int_{-\infty}^{\infty} e^{itx}\, ix\, dG_\alpha(x). \tag{3.3.4}$$

As $G_\alpha(x)$, $0 \leq \alpha \leq 1$ has a uniformly bounded $(1+\delta)$th moment, this family integrates $|x|$ uniformly (cf. Lemma 3.3.1). The continuity of $\frac{\partial}{\partial t}\phi(t,\alpha)$ now follows from (3.3.4), and (b) is proved.

Lemma 3.3.4: Under assumptions (ii), the family $\{G_{\alpha,n}(x), n \geq 1\}$ is relatively compact for every fixed α, that is, every sequence of elements in $\{G_{\alpha,n}(x), n \geq 1\}$ contains a subsequence which converges.

Proof: By Lemma (3.3.1), $|x|^\varepsilon$, $0 \leq \varepsilon < 1+\delta$ is uniformly integrable. Hence the conclusion of the lemma follows from Lemma 3.2.10.

Lemma 3.3.5: If the double sequence of random variables satisfies assumptions (i) - (iii) and (3.3.2), then $\phi(t,\alpha)$, the characteristic function of G_α satisfies the differential equation

$$\frac{\partial}{\partial\alpha}\phi(t,\alpha) = 0 \tag{3.3.5}$$

for $0 < \alpha < 1$, $-\infty < t<\infty$ and has boundary values $\phi(t,o) \equiv 1$.

Proof: Since $G_{o,n} = \delta_o$ for all n, where δ_o is the distribution which has its entire mass at the point zero, the assertion about the boundary values follows from Lemma 3.3.3(a). To derive the differential equation we first show that (3.3.5) is satisfied, if $\frac{\partial}{\partial\alpha}\phi(t,\alpha)$ is changed to the right derivative $\frac{\partial^+}{\partial\alpha}\phi(t, \alpha)$. We have

$$\begin{aligned}\frac{\partial^+}{\partial\alpha}\phi(t,\alpha) &= \lim_{\Delta\to+o} \Delta^{-1} [\phi(t,\alpha+\Delta) - \phi(t,\alpha)] \\ &= \lim_{\Delta\to+o} \Delta^{-1} \lim_{n\to\infty} E[\exp(it\, S_{\alpha+\Delta,n}) - \exp(it\, S_{\alpha,n})] \\ &= \lim_{\Delta\to+o} \Delta^{-1} \lim_{n\to\infty} A_n \qquad (3.3.6)\end{aligned}$$

where

$$A_n = E[\exp(it\, S_{\alpha,n}) \{\exp(it\, T^n_{\alpha\beta_n,\Delta}) - 1\}].$$

We will now prove that this limit value exists from which the existence of $\frac{\partial^+}{\partial\alpha}\phi(t,\alpha)$ follows. By the elementary inequality, there exists a universal constant C such that

$$|H(\lambda)| \le C \min(|\lambda|,\lambda^2)$$

for all real λ, where $H(\lambda) = e^{i\lambda} -1-i\lambda$, it follows that

$$\begin{aligned}|A_n| &= |E[\exp(it\, S_{\alpha,n})\{\exp(itT^n_{\alpha\beta_n,\Delta}) - 1\}]| \\ &= |E[\exp(it\, S_{\alpha,n}) \{it\, T^n_{\alpha\beta_n\,\Delta} + H(t\, T^n_{\alpha\beta_n,\Delta})\}]|\end{aligned}$$

$$\leq |it\, E\{e^{itS_{\alpha,n}}\, E(T^n_{\alpha\beta_n,\Delta}|S_{\alpha,n})\}| + |E\{e^{itS_{\alpha,n}}\, H(tT^n_{\alpha\beta_n,\Delta})\}|$$

$$\leq |t|E|E(T^n_{\alpha\beta_n,\Delta}\ |S_{\alpha,n})| + E|H(t\, T^n_{\alpha\beta_n,\Delta})|$$

$$\leq |t|E|E(T^n_{\alpha\beta_n,\Delta}|S_{\alpha,n})| + C\, E[\min\{|t\, T^n_{\alpha\beta_n,\Delta}|, |tT^n_{\alpha\beta_n,\Delta}|^2\}] \qquad \ldots(3.3.7)$$

Let $\varepsilon>0$. The second term on the right hand side of (3.3.7)

$$= C\int \min\{|tx|\ , t^2x^2\}\ dG^n_{\alpha,\Delta}(x)$$

$$= C \int_{|x|>\varepsilon/|t|} \min\{|tx|, t^2x^2\}\ dG^n_{\alpha,\Delta}(x)$$

$$+ C \int_{|x|\leq\varepsilon/|t|} \min\{|tx|, t^2x^2\}\ dG^n_{\alpha,\Delta}(x)$$

$$\leq C|t| \int_{|x|>\varepsilon/|t|} |x|\ dG^n_{\alpha,\Delta}(x) + C\,\varepsilon\, |t| \int_{|x|\leq\varepsilon/|t|} |x|dG^n_{\alpha.\Delta}(x).$$

$$\leq C|t| \int_{|x|>\varepsilon/|t|} |x|\ dG^n_{\alpha,\Delta}(x) + C\,\varepsilon|t|\ E|T^n_{\alpha\beta_n,\Delta}|$$

Therefore it follows from (3.3.7) that

$$\lim_{\Delta\to+o} \Delta^{-1} \lim_{n\to\infty} |A_n| \leq |t| \lim_{\Delta\to+o} \Delta^{-1} \lim_{n\to\infty} E|E(T^n_{\alpha\beta_n,\Delta}|S_{\alpha,n})|$$

$$+ C|t|\lim_{\Delta\to+o} \Delta^{-1} \lim_{n\to\infty} \int_{|x|>\varepsilon/|t|} |x|dG^n_{\alpha,\Delta}(x)$$

$$+ C\,\varepsilon|t| \lim_{\Delta\to+o} \Delta^{-1} \lim_{n\to\infty} E|T^n_{\alpha\beta_n,\Delta}| \qquad \ldots(3.3.8)$$

The first term on the right side tends to zero by assumption (iii), the second term tends to zero by assumption (iv) and the last term tends to zero since ε is arbitrary. By (3.3.6) it follows that

$$\frac{\partial^+}{\partial\alpha}\phi(t,\alpha) = 0 \; . \qquad (3.3.9)$$

Since the right hand side in (3.3.9) is continuous $\phi(t,\alpha)$ is differentiable with respect to α and is equal to zero. Hence, the lemma follows.

Lemma 3.3.6: The function

$$\phi(t,\alpha) = 1 \qquad (3.3.10)$$

is the unique (complex-valued) solution in the domain $0 \leq \alpha < 1$, $-\infty < t < \infty$ of the differential equation (3.3.5), which satisfies the boundary value condition $\phi(t,0) \equiv 1$ for $-\infty < t < \infty$.

Proof: The general solution of the differential equation is $\phi(t,\alpha) = g(t) + i\,h(t)$. By the boundary condition, we get $g(t) \equiv 1$ and $h(t) \equiv 0$. Since (3.3.5) is a linear homogeneous equation, the solution is unique. This completes the rproof of the lemma.

Now, we give proof of Theorem 3.3.1. From Lemmas 3.3.5, 3.3.6 and from the one-one correspondence between characteristic functions and distribution functions, it follows that

$$\lim_{n\to\infty} \mathcal{L}(S_{\alpha,n}) = \delta_0 \qquad (3.3.11)$$

for $0 \leq \alpha < 1$, when assumptions (i) - (iii) and (3.3.2) are satisfied. Since the family $\{G_\alpha(x), 0 \leq \alpha \leq 1\}$ is continuous by Lemma 3.3.2, we get that (3.3.1) holds for $\alpha = 1$. Next we show that (3.3.2) is actually superfluous, by contradiction. Thus we assume that the double sequence of random variables satisfies assumptions (i) - (iii) and that for some α_o it holds that

$$\lim_{n\to\infty} \mathcal{L}(S_{\alpha_o,n}) \neq \delta_o. \tag{3.3.12}$$

According to Lemma 3.3.4, there is a subsequence $\{n_\mu\}$ such that

$$\lim_{\mu\to\infty} \mathcal{L}(S_{\alpha_o,n_\mu}) = \Lambda \tag{3.3.13}$$

where

$$\Lambda \neq \delta_o. \tag{3.3.14}$$

Again, from Lemma 3.3.4, it follows that we can pick a new subsequence $\{n_\nu\}$ of $\{n_\mu\}$ such that

$$\lim_{\nu\to\infty} G_{\alpha,n_\nu} = G_\alpha \tag{3.3.15}$$

for all rational $\alpha \in [0, 1]$ by a diagonal procedure. As in Lemma 3.3.2, it follows that this family $\{G_\alpha, \alpha$ rational belonging to $[0, 1]\}$ is uniformly continuous in the Lévy metric. Thus we can extend it by continuity to be defined for all $\alpha \in [0, 1]$. Then we get that

$$\lim_{\nu\to\infty} G_{\alpha,n_\nu} = G_\alpha \tag{3.3.16}$$

for $0 \leq \alpha \leq 1$, because for any rational number $r > \alpha$,

$$L(G_{\alpha,n_\nu}, G_\alpha) \leq L(G_\alpha, G_r) + L(G_{\alpha,n_\nu}\ G_{r,n_\nu}) + L(G_{r,n_\nu}, G_r). \quad \ldots(3.3.17)$$

Lemma 3.3.2 and (i) yield that

$$\overline{\lim_{\nu\to\infty}}\, L(G_{\alpha,n_\nu}, G_{r,n_\nu}) \leq \{\chi(r-\alpha)\}^{1/(2+\delta)}. \quad (3.3.18)$$

Thus, by letting $\nu \to \infty$ in (3.3.17), we obtain

$$\overline{\lim_{\nu\to\infty}}\, L(G_{\alpha,n_\nu}, G_\alpha) \leq \inf_{r>\alpha} \{L(G_\alpha, G_r) + \{\chi(r-\alpha)\}^{1/(2+\delta)}\}$$

$$= 0$$

by (3.3.15). Thus (3.3.16) follows. We note that (i) - (iii) are unchanged when we restrict to a subsequence. Since (3.3.16) is the same as (3.3.2), from (3.3.11) it follows that

$$G_\alpha = \lim_{\nu\to\infty} G_{\alpha,n_\nu} = \delta_o. \quad (3.3.19)$$

Thus we get

$$\Lambda = \delta_o$$

which is a controdiction. Thus, under assumptions (i)-(iii),

$$\lim_{n\to\infty} \mathcal{L}(S_{\alpha,n}) = \delta_o \quad (3.3.20)$$

for $0 \leq \alpha \leq 1$. Therefore,

$$(p)\lim_{n\to\infty} S_{\alpha,n} = 0$$

for $0 \leq \alpha \leq 1$ by Theorem 10.1.d of Loéve [28], p. 168, which proves Theorem 3.3.1.

As an application of Theorem 3.3.1, we derive the following weak law of large numbers for double sequences, when the random variables in the same row are stationary and ϕ-mixing.

Theorem 3.3.2: Let $\{Y_{n,i}, i = 1,2,\ldots,\beta_n, n \geq 1\}$ be a double sequence of random variables satisfying the following conditions:

(a) The random variables in the nth row, that is, for fixed n, are stationary and $\phi^{(n)}$-mixing, where

$$\phi_k^{(n)} \leq \psi_k \tag{3.3.21}$$

for all $n \geq 1$, $1 \leq k \leq \beta_n$ and

$$\psi_k \to 0, \text{ as } k \to \infty. \tag{3.3.22}$$

(b) $E(Y_{n,i}) = 0$ for all i and n.

(c) $\overline{\lim_{n\to\infty}} \beta_n^{1+\delta} E|Y_{n,1}|^{1+\delta} = M_1 < \infty$, for some $\delta > 0$.

(d) There exists a sequence of positive integers $\{\gamma_n, n \geq 1\}$ such that

(i) $\gamma_n = o(\beta_n)$, and

(ii) $\psi_{\gamma_n} \to 0$ as $n \to \infty$.

Then

$$(p)\lim_{n\to\infty} S_{1,n} = 0.$$

where $S_{1,n} = Y_{n,1} + \ldots + Y_{n,\beta_n}$.

Proof: We prove this theorem by showing that assumptions (i) - (iv) of Theorem 3.3.1 are satisfied.

Lemma 3.3.7: If the assumptions (a) and (c) are satisfied, then the assumption (ii) of Theorem 3.3.1 will be satisfied.

Proof: By Minkowski inequality,

$$E|S_{\alpha,n} - S_{\gamma,n}|^{1+\delta} = E|\sum_{i=\gamma\beta_n+1}^{\alpha\beta_n} Y_{n,i}|^{1+\delta}$$

$$\leq [\sum_{i=\gamma\beta_n+1}^{\alpha\beta_n} \{E|Y_{n,i}|^{1+\delta}\}^{1/(1+\delta)}]^{1+\delta}$$

$$= E|Y_{n,1}|^{1+\delta}\{(\alpha-\gamma)\beta_n\}^{1+\delta}$$

for some $\delta>0$, by assumption (a). Therefore,

$$\overline{\lim_{n\to\infty}}\, E|S_{\alpha,n} - S_{\gamma,n}|^{1+\delta} \leq (\alpha-\gamma)^{1+\delta} \overline{\lim_{n\to\infty}}\, \beta_n^{1+\delta} E|Y_{n,1}|^{1+\delta}$$

$$= (\alpha-\gamma)^{1+\delta} M_1 , \qquad (3.3.23)$$

by assumption (c). Hence the lemma follows.

Lemma 3.3.8: Under conditions (a) - (d), assumption (iii) of Theorem 3.3.1 will hold.

Proof: For the γ_n in (d), by Lemma 3.2.5, we have

$$E|E(T^n_{\alpha\beta_n,\Delta}|S_{\alpha,n}| \leq E|E(T^n_{\alpha\beta_n+\gamma_n,\Delta}|S_{\alpha,n})|$$

$$+ 2E|T^n_{\alpha\beta_n+\gamma_n,\Delta} - T^n_{\alpha\beta_n,\Delta}|. \qquad (3.3.24)$$

Consider the first term on the right hand side. By Lemma 3.2.2, (a) and (b)

$$E|E(T^n_{\alpha\beta_n+\gamma_n,\Delta}|S_{\alpha,n})| \leq 2\{\phi^{(n)}_{\gamma_n}\}^{\delta/(1+\delta)}\{E|T^n_{\alpha\beta_n+\gamma_n,\Delta}|^{1+\delta}\}^{1/(1+\delta)}$$

$$\leq 2\{\psi_{\gamma_n}\}^{\delta/(1+\delta)}\{\, E|S_{\Delta,n}|^{1+\delta}\}^{1/(1+\delta)} \qquad (3.3.25)$$

by stationarity. By taking $\gamma = 0$ and $\alpha = \Delta$ in the proof of Lemma 3.3.7, the right side of (3.3.25) is

$$\leq 2\{\psi_{\gamma_n}\}^{\delta/(1+\delta)}\ \Delta\ \{\beta_n^{1+\delta}\ E|Y_{n,1}|^{1+\delta}\}^{1/(1+\delta)} .$$

Taking limit as $n\to\infty$ on both sides, we get by (c) and (d),

$$\overline{\lim_{n\to\infty}}\ E|E(T^n_{\alpha\beta_n+\gamma_n,\Delta}|S_{\alpha,n})| = 0. \qquad (3.3.26)$$

Further more, if $R_{\alpha,\gamma_n} = \sum_{i=\alpha\beta_n+1}^{\alpha\beta_n+\gamma_n} Y_{n,i}$,

$$\begin{aligned} E|T^n_{\alpha\beta_n+\gamma_n,\Delta} - T^n_{\alpha\beta_n,\Delta}| &= E|R_{\alpha,\gamma_n} - R_{(\alpha+\Delta),\gamma_n}| \\ &\leq 2E|R_{\alpha,\gamma_n}| \\ &\leq 2\{E|R_{\alpha,\gamma_n}|^{1+\delta}\}^{1/(1+\delta)} \\ &\leq 2\{\gamma_n^{1+\delta}\ E|Y_{n,1}|^{1+\delta}\}^{1/(1+\delta)}\ , \end{aligned}$$

by Minkowski inequality. Assumptions (c) and (d) imply that

$$\lim_{n\to\infty}\ E|T^n_{\alpha\beta_n+\gamma_n,\Delta} - T^n_{\alpha\beta_n,\Delta}| = 0. \qquad (3.3.27)$$

From (3.3.24), (3.3.26) and (3.3.27) the required result follows.

Lemma 3.3.9: If assumptions (a) and (c) are satisfied, then assumption (iv) of Theorem 3.3.1 will be satisfied.

Proof: By (3.3.23), we have

$$\overline{\lim_{n\to\infty}} E|T^n_{\alpha\beta_n,\Delta}|^{1+\delta} = \overline{\lim_{n\to\infty}} E|S_{\alpha+\Delta,n} - S_{\alpha,n}|^{1+\delta}$$

$$\leq \overline{\lim_{n\to\infty}} \Delta^{1+\delta} M_1.$$

Hence

$$\lim_{\Delta\to+o} \Delta^{-1} \overline{\lim_{n\to\infty}} E|T^n_{\alpha\beta_n,\Delta}|^{1+\delta} = 0.$$

Then the lemma follows by an argument similar to that given in (3.3.3).

Thus, it follows from Lemma 3.3.7, Lemma 3.3.8 and assumption (a), that all the conditions of Theorem 3.3.1 are satisfied. Hence the conclusion of Theorem 3.3.2 follows.

3.4 Central Limit Theorem for Double Sequences of Random Variables:

In this section, we deduce the central limit theorem for stationary ϕ-mixing processes from the following theorem of Rosén [40]. We use the same notation as in the previous section.

Theorem 3.4.1: Suppose the double sequence $\{Y_{n,i},\ i=1,\ldots,\beta_n, \beta_n \geq 1\}$ of random variables satisfies the following conditions.

(C1) $\overline{\lim_{n\to\infty}} E(S_{\alpha,n} - S_{\beta,n})^2 \leq \chi(\alpha-\beta)$, $0 \leq \beta < \alpha \leq 1$ for some function $\chi(s)$ which is bounded on $0 \leq s \leq 1$ and tends to zero as s tends to zero.

(C2) There is a function $\rho(\alpha)$, continuous for $0 \le \alpha < 1$, such that for $0 \le \alpha < 1$,

$$\lim_{\Delta\to+0} \Delta^{-1} \overline{\lim_{n\to\infty}} E|E(T^n_{\alpha\beta_n,\Delta}|S_{\alpha,n}) - \Delta\rho(\alpha)\, S_{\alpha,n}| = 0 .$$

(C3) There is a function $\sigma^2(\alpha)$, continuous for $0 \le \alpha < 1$, such that for $0 \le \alpha < 1$,

$$\lim_{\Delta\to+0} \Delta^{-1} \overline{\lim_{n\to\infty}} E|E\{(T^n_{\alpha\beta_n,\Delta})^2 |S_{\alpha,n}\} - \Delta\sigma^2(\alpha)| = 0 .$$

(C4) $\lim_{\Delta\to+0} \Delta^{-1} \overline{\lim_{n\to\infty}} \int_{|x|>\varepsilon} x^2\, dG^n_{\alpha,\Delta}(x) = 0$, $0 \le \alpha < 1$ for every $\varepsilon > 0$.

Then

$$\lim_{n\to\infty} \mathcal{L}(S_{\alpha,n}) = N(0, \psi(\alpha)), \ 0 \le \alpha \le 1 \tag{3.4.1}$$

where

$$\psi(\alpha) = \begin{cases} \int_0^\alpha \sigma^2(s) \exp(2\int_s^\alpha \rho(u)\,du)\,ds, & 0 \le \alpha < 1 \\ \lim_{s\to1} \psi(s) & \text{for } \alpha = 1 \end{cases} \tag{3.4.2}$$

and $N(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2.

Proof: For proof of this theorem, we refer to Rosén [40].

The following theorem is a particular version of the above theorem for stationary ϕ-mixing processes.

Lemma 3.4.1: If assumptions (i), (ii), (iv) and (v) are satisfied, then

$$\overline{\lim_{n\to\infty}} E(S_{\alpha,n} - S_{\varepsilon,n})^2 \le \chi(\alpha-\varepsilon),\ 0 \le \varepsilon < \alpha \le 1 \qquad (3.4.6)$$

for some function $\chi(s)$ which is bounded on $0 \le s \le 1$ and tends to zero as $s \to 0$.

Proof: We have,

$$\begin{aligned} E(S_{\alpha,n} - S_{\varepsilon,n})^2 &= E\Big(\sum_{i=\varepsilon\beta_n+1}^{\alpha\beta_n} Y_{n,i}\Big)^2 \\ &= \sum_{i=\varepsilon\beta_n+1}^{\alpha\beta_n} E(Y_{n,i}^2) + 2\sum_{i=\varepsilon\beta_n+1}^{\alpha\beta_n}\ \sum_{j=i+1}^{\alpha\beta_n} E(Y_{n,i}Y_{n,j}) \\ &\le (\alpha-\varepsilon)\beta_n E(Y_{n,1}^2) \\ &\quad + 2\sum_{i=\varepsilon\beta_n+1}^{\alpha\beta_n}\ \sum_{j=i+1}^{\alpha\beta_n} |E(Y_{n,1}\,Y_{n,j-i+1})| \\ &\le (\alpha-\varepsilon)\beta_n E(Y_{n,1}^2) \\ &\quad + 2\sum_{i=\varepsilon\beta_n+1}^{\alpha\beta_n}\ \sum_{j=i+1}^{\alpha\beta_n} \{\phi_{j-i}^{(n)}\}^{1/2} E(Y_{n,1}^2) \end{aligned} \qquad \ldots(3.4.7)$$

by Lemma 3.2.3 and assumption (i) and (ii). Therefore, by (3.4.3) we get

$$E(S_{\alpha,n} - S_{\varepsilon,n})^2 \le (\alpha-\varepsilon)\,\beta_n\,E(Y_{n,1}^2) + 2\,E(Y_{n,1}^2) \sum_{i=\varepsilon\beta_n+1}^{\alpha\beta_n}\ \sum_{k=1}^{\alpha\beta_n-i} \psi_k^{1/2}$$

$$\leq E(Y_{n,1})^2 \{(\alpha-\varepsilon)\beta_n + 2 \sum_{i=\varepsilon\beta_n+1}^{\alpha\beta_n} \sum_{k=1}^{\infty} \{O(k^{-\theta})\}^{1/2}\} \quad (3.4.8)$$

by assumption (v). That is,

$$E(S_{\alpha,n} - S_{\varepsilon,n})^2 \leq (\alpha-\varepsilon)\ \beta_n\ E(Y^2_{n,1})\ (1+2M_3) \quad (3.4.9)$$

where

$$\sum_{k=1}^{\infty} \{O(k^{-\theta})\}^{1/2} = M_3 < \infty, \quad (3.4.10)$$

by assumption v). Note that (iv) implies that $\overline{\lim_n}\ \beta_n\ E(Y^2_{n,1}) < \infty$. Therefore, by taking limit as $n \to \infty$ on both sides of (3.4.9), the lemma follows.

Lemma 3.4.2: If assumptions (i), (ii), (iv) and (v) are satisfied, then

$$\lim_{\Delta\to+0} \Delta^{-1} \overline{\lim_{n\to\infty}}\ E|E(T^n_{\alpha\beta_n,\Delta}|S_{\alpha,n})| = 0,\ 0 \leq \alpha < 1. \quad (3.4.11)$$

Proof: By Lemma 3.2.2 and assumption (i) we have

$$\begin{aligned} E|E(Y_{n,\alpha\beta_n+i}|S_{\alpha,n})| &\leq 2\{\phi_i^{(n)}\ E(Y^2_{n,\alpha\beta_n+i})\}^{1/2} \\ &\leq 2\{\psi_i\ E(Y^2_{n,1})\}^{1/2}, \end{aligned}$$

by assumption (i). Therefore, we get

$$\begin{aligned} E|E(T^n_{\alpha\beta_n,\Delta}|S_{\alpha,n})| &\leq 2\{E(Y^2_{n,1})\}^{1/2} \sum_{i=1}^{\Delta\beta_n} \psi_i^{1/2} \\ &\leq 2\{E(Y^2_{n,1})\}^{1/2} \sum_{i=1}^{\infty} \psi_i^{1/2} \\ &= 2M_3\{E(Y^2_{n,1})\}^{1/2} \quad (3.4.12) \end{aligned}$$

by assumption (v) and (3.4.10). Therefore,

$$E|E(T^n_{\alpha\beta_n,\Delta}|S_{\alpha,n})| \le 2M_3\,\beta_n^{-1/2}\,\{\beta_n\,E(Y^2_{n,1})\}^{1/2}.$$

By taking limit supremum on both sides as $n \to \infty$, the lemma follows from assumption (iv) and the fact that $\beta_n \to \infty$ as $n \to \infty$.

Lemma 3.4.3: Under assumptions (i), (ii), (iv) and (v) and (3.4.5) $\nu(\alpha)$ is non-decreasing and satisfies Lipschitz condition on $0 \le \alpha \le 1$.

Proof: By assumption (ii), we have

$$E(S^2_{\alpha+\Delta,n}) = E(S^2_{\alpha,n}) + E(T^n_{\alpha\beta_n,\Delta})^2 + 2E(T^n_{\alpha\beta_n,\Delta}\cdot S_{\alpha,n}).$$

Now,taking limits as $n \to \infty$, on both sides we get

$$\nu(\alpha+\Delta) - \nu(\alpha) = \lim_{n\to\infty}\,[E(T^n_{\alpha\beta_n,\Delta})^2 + 2E(T^n_{\alpha\beta_n,\Delta}\cdot S_{\alpha,n})] \tag{3.4.13}$$

by (3.4.5). Consider the second term on the right side of (3.4.13). Then,

$$|E(T^n_{\alpha\beta,\Delta}\cdot S_{\alpha,n})| \le \left|E\{S_{\alpha,n}(\sum_{i=\alpha\beta_n+1}^{p_n} Y_{n,i})\}\right| + E\{S_{\alpha,n}(\sum_{i=p_n+1}^{(\alpha+\Delta)\beta_n} Y_{n,i})\} \tag{3.4.14}$$

where $p_n = \alpha\beta_n + \Delta\,\gamma_n$ and $\{\gamma_n\}$ is a sequence of non-negttive integers such that

(i) $\gamma_n \to \infty$ as $n \to \infty$, (3.4.15)

and (ii) $\gamma_n = o(\beta_n)$, (3.4.16)

Using Schwartz inequality for the first term and Lemma 3.2.3 to the second term of (3.4.14), by assumption (ii), it follows that,

$$|E(T^n_{\alpha\beta_n,\Delta}\cdot S_{\alpha,n})| \leq \{E(S^2_{\alpha,n}).E(\sum_{i=\alpha\beta_n+1}^{p_n} Y_{n,i})^2\}^{1/2}$$

$$+ 2\{\phi^{(n)}_{\Delta\gamma_n} E(S^2_{\alpha,n}).\ E(\sum_{i=p_n+1}^{(\alpha+\Delta)\beta_n} Y_{n,i})^2\}^{1/2}$$

$$\leq \{E(S^2_{\alpha,n})E(\sum_{i=1}^{\Delta\gamma_n} Y_{n,i})^2\}^{1/2}$$

$$+ 2\{\psi_{\Delta\gamma_n} E(S^2_{\alpha,n})\ E(\sum_{i=1}^{\Delta(\beta_n-\gamma_n)} Y_{n,i})^2\}^{1/2}. \qquad \ldots(3.4.17)$$

by assumption (i). By methods similar to those given in Lemma 3.4.1, we get

$$|E(T^n_{\alpha\beta_n,\Delta}\cdot S_{\alpha,n})| \leq E(S^2_{\alpha,n})(1+2M_3)\Delta\gamma_n\ E(Y^2_{n,1})\}^{1/2}$$

$$+ 2\{\psi_{\Delta\gamma_n} E(S^2_{\alpha,n})\ (1+2M_3)\Delta(\beta_n-\gamma_n)E(Y^2_{n,1})\}^{1/2}.$$

Hence, by taking limit as $n\to\infty$ on both sides, the second term on the right hand side of (3.4.13) tends to zero, by assumptions (iv), (v), (3.4.5), (3.4.15) and (3.4.16). Therefore (3.4.13) reduces to

$$\nu(\alpha+\Delta) - \nu(\alpha) = \lim_{n\to\infty} E(T^n_{\alpha\beta_n,\Delta})^2 \qquad (3.4.18)$$

$$\leq \lim_{n\to\infty} \{(1+2M_3)\Delta\ \beta_n\ E(Y^2_{n,1})\}$$

by Lemma 3.4.1. Then by assumption (iv), we get

$$\nu(\alpha+\Delta) - \nu(\alpha) \leq (1+2M_3)\Delta M_1 , \tag{3.4.19}$$

where M_1 is a finite constant. From (3.4.18), it follows that $\nu(\alpha)$ is non-decreasing and the lemma follows from (3.4.19). Since $\nu(\alpha)$ is non-decreasing and satisfies Lipschitz condition, $\nu'(\alpha)$, the derivative of $\nu(\alpha)$ with respect to α exists.

Lemma 3.4.4: Under assumptions of the previous lemma

$$\lim_{\Delta\to+0} \Delta^{-1} \overline{\lim_{n\to\infty}} E|E\{(T^n_{\alpha\beta_n,\Delta})^2 |S_{\alpha,n}\} - \Delta\nu'(\alpha)| = 0 \tag{3.4.20}$$

for $0 \leq \alpha \leq 1$, where $\nu'(\alpha)$ is the derivative of $\nu(\alpha)$ with respect to α.

Proof: Consider,

$$E|E\{(T^n_{\alpha\beta_n,\Delta})^2 |S_{\alpha,n}\} - \Delta\nu'(\alpha)|$$

$$\leq E|E\{(T^n_{\alpha\beta_n,\Delta})^2 | S_{\alpha,n}\} - E(T^n_{\alpha\beta_n,\Delta})^2|$$

$$+ E|E (T^n_{\alpha\beta_n,\Delta})^2 - \Delta\nu'(\alpha)| . \tag{3.4.21}$$

Since, by (3.4.19),

$$\nu(\alpha+\Delta) - \nu(\alpha) = \lim_{n\to\infty} E(T^n_{\alpha\beta_n,\Delta})^2$$

$$\leq (1+2M_3)\Delta M_1$$

An application of dominated convergence theorem shows that

$$\overline{\lim_{\Delta\to+0}}\ \Delta^{-1}\ \overline{\lim_{n\to\infty}}\ E|E(T^{n}_{\alpha\beta_n,\Delta})^2 - \Delta\nu'(\alpha)| = 0 . \qquad (3.4.22)$$

If we take $X_{n,i} = \beta_n^{1/2}\ Y_{n,i}$ in Lemma 3.2.7, all the conditions of Lemma 3.2.7 will be satisfied and hence we obtain that

$$E|E\{(T^{n}_{\alpha\beta_n,\Delta})^2\ |S_{\alpha,n}\} - E(T^{n}_{\alpha\beta_n,\Delta})^2|$$
$$\leq C(\Delta^2+\Delta^{1/2})\ \beta_n^{-\ell}\ \{\beta_n^{1+\varepsilon/2}\ E|Y_{n,1}|^{2+\varepsilon}\}^{2/(2+\varepsilon)}$$

for all large n and $\ell>0$. Taking limits as $n\to\infty$ on both sides by, assumption (iv) the right side tends to zero and hence the lemma follows from (3.4.21) and (3.4.22).

Lemma 3.4.5: Under assumptions of the theorem 3.4.2,

$$\lim_{\Delta\to+0}\ \Delta^{-1}\ \overline{\lim_{n\to\infty}} \int_{|x|>\varepsilon} x^2\ dG^{n}_{\alpha,\Delta}(x) = 0 , \quad 0 \leq \alpha < 1 \qquad (3.4.23)$$

for every $\varepsilon>0$.

Proof: When assumptions of the Theorem 3.4.2 are satisfied, if we take $X_{n,i} = \beta_n^{1/2}\ Y_{n,i}$ in Lemma 3.2.7, it follows that

$$E|\sum_{i=1}^{\Delta\beta_n} Y_{n,i}|^{2+\tau} \leq \Delta^{1+\tau/2}\ M_2 \qquad (3.4.24)$$

from (3.2.49) for all $\beta_n > N$. Therefore

$$E|T^{n}_{\alpha\beta_n,\Delta}|^{2+\tau} = E|\sum_{i=1}^{\Delta\beta_n} Y_{n,i}|^{2+\tau} \leq \Delta^{1+\tau/2}\ M_2 .$$

We have

$$\int_{|x|>\varepsilon} x^2 \, dG^n_{\alpha,\bar{\Delta}}(x) \leq \varepsilon^{-\tau} \int_{|x|>\varepsilon} |x|^{2+\tau} \, dG^n_{\alpha,\bar{\Delta}}(x)$$

$$\leq \varepsilon^{-\tau} E|T^n_{\alpha\beta_n,\Delta}|^{2+\tau}$$

$$\leq \varepsilon^{-\tau}\Delta^{1+\tau/2} M_2. \qquad (3.4.25)$$

for all $\beta_n > N$ where $\tau > 0$. Hence the lemma follows.

Proof of Theorem 3.4.2: When the assumptions of the theorem and (3.4.5) are satisfied, it follows from Lemmas 3.4.1, 3.4.2, 3.4.4 and 3.4.5 that all the four conditions of Theorem 3.4.1 are satisfied. Hence the theorem follows from Theorem 3.4.1. Now, we will show that (3.4.5) is actually superfluous. Let

$$\nu_n(\alpha) = E\left(\sum_1^{\alpha\beta_n} Y_{n,i}\right)^2 \qquad (3.4.26)$$

for $0 \leq \alpha \leq 1$. The functions $\nu_n(\alpha)$ satisfy:

$$\nu_n(0) = 0, \quad \nu_n(1) = 1 \qquad (3.4.27)$$

and

$$\nu_n(\alpha+\Delta) - \nu_n(\alpha) = E(S^2_{\alpha+\Delta,n} - S^2_{\alpha,n})$$

$$= E(T^n_{\alpha\beta_n,\Delta})^2 + 2E(S_{\alpha,n} T^n_{\alpha\beta_n,\Delta})$$

$$\leq (2(\Delta\alpha)^{1/2} + \Delta)\, \beta_n E(Y^2_{n,1}) \qquad (3.4.28)$$

by stationarity, Schwartz inequality and arguments in Lemma 3.4.1. From (3.4.27) and (3.4.28), it follows that the sequence $\{\nu_n(\alpha), n \geq 1\}$ contains a sub-sequence which converges, i.e., there exists $\{n_k\}$ such that

$$\lim_{k\to\infty} \nu_{n_k}(\alpha) = \nu(\alpha), \quad 0 \leq \alpha \leq 1. \tag{3.4.29}$$

(3.4.5) is satisfied for this sub-sequence. By arguments similar to those given in the case of weak law of large numbers the theorem follows.

3.5 Regularity Conditions:

Let $\{X_n, n \geq 1\}$ be a sequence of stationary ϕ-mixing random variables. We denote by $p_n(x_1,\ldots,x_n;\theta)$ the n-dimensional joint density function which depends on a single unknown parameter θ. We have

$$\begin{aligned}\log p_n(x_1,\ldots,x_n;\theta) &= \sum_{i=1}^{n} \log p(x_i;\theta|x_{i-1},\ldots,x_1)\\ &= \log p(x_1\ldots,x_{\ell_n};\theta)\\ &\quad + \sum_{i=\ell_n+1}^{n} \log p(x_i;\theta\,|x_{i-1},\ldots,x_1)\end{aligned} \qquad \ldots(3.5.1)$$

where $\{\ell_n\}$ is a sequence of non-negative integers such that

$$\ell_n \to C < \infty, \text{ as } n \to \infty. \tag{3.5.2}$$

We shall define the approximate log-likelihood function given the observations to be

$$\log \tilde{p}_n(\theta) = \sum_{i=\ell_n+1}^{n} \log p(x_i;\theta\,|x_{i-1},\ldots,x_{i-\ell_n}), \tag{3.5.3}$$

where $\tilde{p}_n(\theta) = \tilde{p}_n(x_1,\ldots,\theta)$. Since $\{X_n, n \geq 1\}$ is stationary, (3.5.3) can be written as

$$\log \tilde{p}_n(\theta) = \sum_{i=1}^{\beta_n} \log p_{i,n}(\theta) \tag{3.5.4}$$

where

$$\beta_n = n - \ell_n \tag{3.5.5}$$

and

$$p_{i,n}(\theta) = p(x_{i+\ell_n};\theta | x_{i+\ell_n-1},\ldots,x_i). \tag{3.5.6}$$

Definition 3.5.1: An estimator $\hat{\theta}_n \equiv \hat{\theta}_n(x_1,\ldots,x_n)$ of the unknown parameter θ, which maximizes $\log \tilde{p}_n(\theta)$ in a neighbourhood of the true parameter value θ_o, is said to be an approximate maximum likelihood estimator (AMLE) of θ.

Suppose the following regularity conditions are satisfied for all n.

(A1) The parameter space Θ is an open interval on the real line and thr true parameter value θ_o is an interior point of Θ.

(A2) $\frac{\partial^i}{\partial\theta^i} \log p_n(\theta)$, $i = 1,2$ exist for all $\theta\varepsilon\Theta$ and are continuous in θ with probability 1, where

$p_n(\theta) = p(x_n;\theta|x_{n-1},\ldots, x_1)$.

For all $\theta\varepsilon\Theta$ and for almost all $x_1,\ldots,x_{n-1}$,

(A3) $E_\theta[\frac{\partial}{\partial\theta} \log p_n(\theta) \mid x_1,\ldots,x_{n-1}] = 0$

(A4)(i) $-E_\theta[\frac{\partial^2}{\partial\theta^2} \log p_n(\theta)|x_1,\ldots,x_{n-1}]$

$$= E_\theta[\{\frac{\partial}{\partial\theta} \log p_n(\theta)\}^2|x_1,\ldots x_{n-1}]$$

(ii)
$$\begin{aligned} i(\theta) &= -\lim_{n\to\infty} n^{-1} E_\theta[\frac{\partial^2}{\partial\theta^2} \log p(x_1,\ldots,x_n)] \\ &= \lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} \sigma_k^2(\theta) \\ &< \infty \end{aligned}$$

where $\sigma_n^2(\theta) = E_\theta[\frac{\partial}{\partial\theta} \log p_n(\theta)]^2$

and $i(\theta)$ is positive, and

(iii) for the sequence $\{\ell_n\}$ in (3.5.2),

$$\lim_{n\to\infty} E_\theta[\frac{\partial}{\partial\theta} \log p(x_{\ell_n+1}| x_1,\ldots,x_{\ell_n})]^2 = i(\theta).$$

(A5) For every $\theta \in \Theta$, there exists a neighbourhood $V(\theta)$ of θ such that, for $\theta' \in V(\theta)$

$$\begin{aligned} &|\frac{\partial^2}{\partial\theta^2} \log p_n(\theta') - \frac{\partial^2}{\partial\theta^2} \log p_n(\theta)| \\ &\qquad \le |\theta-\theta'| \; G(x_1,x_2,\ldots,x_n) \end{aligned}$$

where $0 \le G(x_1,\ldots,x_n)$ and $\lim_{n\to\infty} E_\theta[G(X_1,\ldots,X_n)]^{1+\delta} < \infty$,

for some $\delta > 0$, independent of θ.

(A6) For all $\theta \in \Theta$,

$$\lim_{n\to\infty} E_\theta|\frac{\partial^2}{\partial\theta^2} \log p_n(\theta)|^{1+\delta} < \infty$$

for some $\delta > 0$, independent of θ.

Hereafter all statements which involve convergence of random variables are valid with respect to the true probability measure P_{θ_o} unless otherwise specified. We denote $\log \tilde{p}_n(\theta)$ by $\psi_n(\theta)$ and $\log p_{i,n}(\theta)$ by $\psi_{i,n}(\theta)$. $\psi_n'(\theta)$ and

$\psi'_{i,n}(\theta)$ denote the derivatives of $\psi_n(\theta)$ and $\psi_{i,n}(\theta)$ respectively with respect to θ and $\psi''_n(\theta)$ and $\psi''_{i,n}(\theta)$ denote the second derivatives of $\psi_n(\theta)$ and $\psi_{i,n}(\theta)$ with respect to θ.

3.6 Asymptotic Properties of AMLE:

In this section, we obtain weak consistency, asymptotic normality and first-order efficiency of AMLE.

Theorem 3.6.1: When the conditions of Section 3.5 are satisfied, the likelihood equation has a root, which is consistent.

Proof: Expanding $\psi'_n(\theta) = \frac{\partial}{\partial\theta} \log \tilde{p}_n(\theta)$ about the true value θ_o of the parameter θ, we get

$$\psi'_n(\theta) = \psi'_n(\theta_o) + (\theta-\theta_o)\,\psi''_n(\theta_o) + (\theta-\theta_o)\{\psi''_n(\theta')-\psi''_n(\theta_o)\} \quad \ldots(3.6.1)$$

where $\theta' = \theta_o + r(\theta-\theta_o)$ and $|r| < 1$. By (3.5.4) the right side reduces to

$$\sum_{i=1}^{\beta_n} \psi'_{i,n}(\theta_o) + (\theta-\theta_o) \sum_{i=1}^{\beta_n} \psi''_{i,n}(\theta_o)$$
$$+ (\theta-\theta_o) \sum_{i=1}^{\beta_n} \{\psi''_{i,n}(\theta') - \psi''_{i,n}(\theta_o) \quad (3.6.2)$$

where $\psi'_{i,n}(\theta)$, $\psi''_{i,n}(\theta)$ and β_n are as defined in Section 3.5. It follows from assumption (A5) that there exists a neighbourhood $V(\theta_o)$ such that for $\theta' \in V(\theta_o)$,

$$|\psi''_{i,n}(\theta') - \psi''_{i,n}(\theta_o)| \leq |r(\theta-\theta_o)| G(x_i,\ldots,x_{i+\ell_n-1}) \quad (3.6.3)$$

where $0 \leq G(x_i,\ldots,x_{i+\ell_n-1})$ is such that

$$\overline{\lim_{n\to\infty}} E_{\theta_o} |G(X_1,\ldots,X_{\ell_n})|^{1+\delta} = M < \infty. \quad (3.6.4)$$

Therefore from (3.6.1) - (3.6.3) we get

$$\beta_n^{-1} \psi'_n(\theta) = B_o + (\theta-\theta_o) B_1 + (\theta-\theta_o)^2 r_n B_2 \quad (3.6.5)$$

where $|r_n| < 1$ for all n and

$$B_o = \beta_n^{-1} \sum_{i=1}^{\beta_n} \psi'_{i,n}(\theta_o), \quad (3.6.6)$$

$$B_1 = \beta_n^{-1} \sum_{i=1}^{\beta_n} \psi''_{i,n}(\theta_o), \quad (3.6.7)$$

and

$$B_2 = \beta_n^{-1} \sum_{i=1}^{\beta_n} G(x_i,\ldots,x_{i+\ell_n}). \quad (3.6.8)$$

(3.6.6) - (3.6.8) can be written as

$$B_o = \sum_{i=1}^{\beta_n} Y_{n,i} + E_{\theta_o} \psi'_{i,n}(\theta_o) \quad , \quad (3.6.9)$$

$$B_1 = \sum_{i=1}^{\beta_n} Z_{n,i} + E_{\theta_o} \psi''_{1,n}(\theta_o), \quad (3.6.10)$$

and

$$B_2 = \sum_{i=1}^{\beta_n} \eta_{n,i} + E_{\theta_o} G(x_1,\ldots,x_{\ell_n}) \quad , \quad (3.6.11)$$

where

$$Y_{n,i} = \beta_n^{-1} \{\psi'_{i,n}(\theta_o) - E_{\theta_o} \psi'_{i,n}(\theta_o)\} \quad , \quad (3.6.12)$$

$$Z_{n,i} = \beta_n^{-1} \{\psi''_{i,n}(\theta_o) - E_{\theta_o} \psi''_{i,n}(\theta_o)\} \quad , \quad (3.6.13)$$

and

$$\eta_{n,i} = \beta_n^{-1}\{G(x_i,\ldots,x_{i+\ell_n}) - E_{\theta_o} G(x_i,\ldots,x_{i+\ell_n})\} \quad \ldots(3.6.14)$$

Since $\{X_n, n \geq 1\}$ is stationary, ϕ-mixing sequence of random variables, by Lemma 3.2.1, it follows that for fixed n, $\{Y_{n,i}, i \geq 1\}$; $\{Z_{n,i}, i \geq 1\}$ and $\{\eta_{n,i}, i \geq 1\}$ are also stationary and $\phi^{(n)}$-mixing sequences of random variables, where

$$\phi_k^{(n)} \leq \phi_{k-\ell_n} \tag{3.6.15}$$

for $k > \ell_n$ and 1 otherwise. Furthermore by assumptions (A4), (A6), (3.5.2) and (3.6.4) it follows that the assumptions of Theorem 3.3.2 are satisfied by $\{Y_{n,i}\}$; $\{Z_{n,i}\}$; and $\{\eta_{n,i}\}$ with $\gamma_n = \beta_n^\lambda$, $0 < \lambda < 1$. Hence by Theorem 3.3.2, we get

$$(p)\lim_{n\to\infty} \sum_{i=1}^{\beta_n} Y_{n,i} = 0 \tag{3.6.16}$$

$$(p)\lim_{n\to\infty} \sum_{i=1}^{\beta_n} Z_{n,i} = 0 , \tag{3.6.17}$$

and

$$(p)\lim_{n\to\infty} \sum_{i=1}^{\beta_n} \eta_{n,i} = 0 . \tag{3.6.18}$$

By assumptions (A3) and (A4) we have

$$\lim_{n\to\infty} E_{\theta_o} \psi'_{1,n}(\theta_o) = 0 ,$$

$$\lim_{n\to\infty} E_{\theta_o} \psi''_{1,n}(\theta_o) = - i(\theta_o)$$

and

$$\lim_{n\to\infty} E_{\theta_o} [G(X_1,\ldots,X_{\ell_n})] = M < \infty.$$

Therefore, by (3.6.9) - (3.6.11) and (3.6.16) - (3.6.18) we get

$$\underset{n\to\infty}{(p)\lim} B_o = 0 \qquad (3.6.19)$$

$$\underset{n\to\infty}{(p)\lim} B_1 = -i(\theta_o) \qquad (3.6.20)$$

and

$$\underset{n\to\infty}{(p)\lim} B_2 = M . \qquad (3.6.21)$$

Now, let $\varepsilon > 0$. Choose $\tau > 0$ such that $\{\theta:\theta_o-\tau \le \theta \le \theta_o+\tau\}$ is contained in $V(\theta_o)$. Let $S_1 = \{\underset{\sim}{X} : |B_o| > \tau^2\}$; $S_2 = \{\underset{\sim}{X} : |B_1|+ i(\theta_o)| > \tau\}$; $S_3 = \{\underset{\sim}{X} : |B_2| > M+1\}$ and $S' = S_1 \cup S_2 \cup S_3$ where S' is the complement of the set S.

Since B_o, B_1 and B_2 converge in probability, there exists an integer $n_o(\tau,\varepsilon) > 0$ such that for $n > n_o(\tau,\varepsilon)$, $P_{\theta_o}(S_1) \le \varepsilon/3$, $P_{\theta_o}(S_2) \le \varepsilon/3$ and $P_{\theta_o}(S_3) \le \varepsilon/3$ so that $P_{\theta_o}(S') \le \varepsilon$ and $P_{\theta_o}(S) > 1-\varepsilon$. For $\theta = \theta_o \pm \tau$, we have, by (3.6.5) that

$$\beta_n^{-1} \psi_n'(\theta) = \beta_n^{-1} \frac{\partial}{\partial\theta} \log \tilde{p}_n(\theta)$$

$$= B_o \pm B_1\tau + B_2 r_n \tau^2 . \qquad (3.6.22)$$

For $n > n_o(\tau,\varepsilon)$, with probability exceeding $1-\varepsilon$, $|B_o+\tau^2 r_n B_2| < \tau^2 (M+2)$ and $-B_1\tau > (i(\theta_o)-\tau)\tau$. So if we choose $\tau < i(\theta_o) (M+3)^{-1}$ the sign of $B_o \pm B_1\tau+\tau^2 r_n B_2$ depends on that of $-B_1\tau$ with probability greater than $1-\varepsilon$ for $n > n_o(\tau,\varepsilon)$. Therefore,

$$\frac{\partial}{\partial\theta} \log \tilde{p}_n(\theta) > 0 \quad \text{for } \theta = \theta_o - \tau \text{ and}$$

$$\frac{\partial}{\partial\theta} \log \tilde{p}_n(\theta) < 0 \quad \text{for } \theta = \theta_o + \tau$$

with probability greater than 1-ε for $n > n_o(\tau,\varepsilon)$. Since $\frac{\partial}{\partial\theta} \log \tilde{p}_n(\theta)$ is continuous, there exists a value $\hat{\theta}_n$ of θ such that

$$\frac{\partial}{\partial\theta} \log \tilde{p}_n(\hat{\theta}_n) = 0$$

for $\hat{\theta}_n$ in $[\theta_o-\tau, \theta_o+\tau]$ with probability greater than 1-ε and for $n > n_o(\tau,\varepsilon)$. Hence under assumptions of Section 3.5 the approximate likelihood equation has a consistent root.

Suppose the following two conditions are satisfied in addition to the assumptions of the previous section.

(A7) For all $\theta \in \Theta$,

$$\lim_{n\to\infty} E_\theta \left|\frac{\partial}{\partial\theta} \log p_n(\theta)\right|^{2+\delta} = K(\theta) < \infty$$

for $0 < \delta \leq 1$, δ independent of θ.

(A8) The mixing coefficient ϕ_n satisfies the condition

$$\phi_n = O(n^{-\theta}) \qquad (3.6.23)$$

for $\theta > 1+2/\delta$ where δ is as in (A7).

Theorem 3.6.2: Under the assumptions of Section 3.5, (A7) and (A8) the approximate maximum likelihood estimator is asymptotically normal.

Proof: If $\hat{\theta}_n$ is an AMLE, then by the previous theorem it is consistent and from (3.6.5) we have

$$\beta_n^{-1} \frac{\partial}{\partial\theta} \log \tilde{p}_n(\hat{\theta}_n) = B_o + (\hat{\theta}_n-\theta_o)B_1 + (\hat{\theta}_n-\theta_o)^2 r_n B_2 = 0 \qquad \ldots(3.6.24)$$

where B_o, B_1 and B_2 are as defined in (3.6.6) - (3.6.8). Therefore,

$$\{\beta_n\ i(\theta_o)\}^{1/2}\ (\hat{\theta}_n - \theta_o) = \xi_n\ \eta_n^{-1}\ , \tag{3.6.25}$$

where

$$\xi_n = \beta_n^{1/2}\ \{i(\theta_o)\}^{-1/2}\ B_o, \tag{3.6.26}$$

and

$$\eta_n = -\{i(\theta_o)\}^{-1}[B_1 + (\hat{\theta}_n - \theta_o) r_n B_2]\ . \tag{3.6.27}$$

By (3.6.20) and (3.6.21), $(p)\lim\limits_{n\to\infty} B_1 = -i(\theta_o)$ and $(p)\lim\limits_{n\to\infty} B_2 = M$. Since $\hat{\theta}_n$ is consistent, $(p)\lim\limits_{n\to\infty} (\hat{\theta}_o - \theta_o) = 0$. Hence,

$$(p)\lim_{n\to\infty} \eta_n = 1. \tag{3.6.28}$$

Now, consider ξ_n. By (3.6.6),

$$\xi_n = \{\beta_n\ i(\theta_o)\}^{-1/2} \sum_{i=1}^{\beta_n} \psi'_{i,n}(\theta_o)$$

$$= [\{i(\theta_o)\}^{-1}\ E(S^2_{1,n})]^{1/2} \sum_{i=1}^{\beta_n} Y_{n,i} \tag{3.6.29}$$

where

$$Y_{n,i} = \{\beta_n\ E(S^2_{1,n})\}^{-1/2}\ \psi'_{i,n}(\theta_o), \tag{3.6.30}$$

$$S_{1,n} = \beta_n^{-1/2} \sum_{i=1}^{\beta_n} \psi'_{i,n}(\theta_o). \tag{3.6.31}$$

We have

$$E(S^2_{1,n}) = \beta_n^{-1}\ E[\sum_{i=1}^{\beta_n} \psi'_{i,n}(\theta_o)]^2$$

$$= E[\psi'_{1,n}(\theta_o)]^2 + 2\beta_n^{-1} \sum_{i=1}^{\beta_n} (\beta_n - i)[E\psi'_{1,n}(\theta_o)\psi'_{1+i,n}(\theta_o)]$$

$$= E[\psi'_{1,n}(\theta_o)]^2 + 2 \sum_{i=1}^{\beta_n} [E\psi'_{1,n}(\theta_o)\psi'_{1+i,n}(\theta_o)] + p_n$$

where

$$|p_n| \leq \beta_n^{-1} \sum_{i=1}^{\beta_n} i|E\psi'_{1,n}(\theta_o)\psi'_{1+i,n}(\theta_o)|.$$

Since $E_{\theta_o} \psi'_{i,n}(\theta_o) = 0$ for all n and i and since $\{\psi'_{i,n}(\theta_o),\ i \geq 1\}$ are stationary, $\phi^{(n)}$-mixing sequence of random variables for fixed n, we get, by Lemma 3.2.3, that

$$|p_n| \leq 2\beta_n^{-1} E[\psi'_{1,n}(\theta_o)]^2 \sum_{i=1}^{\beta_n} i\{\phi_i^{(n)}\}^{1/2}.$$

By Lemma 3.2.1, $\phi_i^{(n)} \leq \phi_{i-\ell_n}$ for all i. Therefore

$$|p_n| \leq 2E[\psi'_{1,n}(\theta_o)]^2 \beta_n^{-1} \sum_{i=1}^{\beta_n} i\, \phi_{i-\ell_n}^{1/2}.$$

Taking limits on both sides as $n \to \infty$, by (3.6.23) and assumption (A4), we get that

$$\lim_{n\to\infty} p_n = 0$$

so that

$$q(\theta_o) = \lim_{n\to\infty} E(S_{1,n}^2) = \lim_{n\to\infty} E[\psi'_{1,n}(\theta_o)]^2$$

$$+ 2 \lim_{n\to\infty} \sum_{i=1}^{\beta_n} [E\psi'_{1,n}(\theta_o)\psi'_{1+1,n}(\theta_o)].$$

Therefore

$$q(\theta_o) = i(\theta_o) + g(\theta_o) \tag{3.6.32}$$

where $g(\theta_o) = 2 \lim_{n\to\infty} \sum_{i=1}^{\beta_n} E[\psi'_{1,n}(\theta_o)\psi'_{1+i,n}(\theta_o)]$. Since $\{X_n, n \geq 1\}$ is stationary and ϕ-mixing sequence of random variables satisfying assumption (A8), by Lemma 3.2.1, we get that $\{Y_{n,i}, i \geq 1\}$ for fixed n is stationary $\phi^{(n)}$-mixing, where $\phi_k^{(n)} \leq \phi_{k-\ell_n}$ for $k > \ell_n$ and 1 otherwise. Hence, the double sequence of random variables $\{Y_{n,i}\}$ satisfies assumption (i) of Theorem 3.4.2. $E(\sum_{i=1}^{\beta_n} Y_{n,i})^2 = 1$ by (3.6.30) which is assumption (iii) of Theorem 3.4.2. Assumption (iv) follows from regularity condition (A7). Thus all the conditions of Theorem 3.4.2 are satisfied by $\{Y_{n,i}\}$. Hence it follows that

$$(\mathcal{L})\lim_{n\to\infty} \sum_{i=1}^{\beta_n} Y_{n,i} = N(0,1) \tag{3.6.33}$$

Since, $\lim_{n\to\infty} E(S_{1,n}^2) = q(\theta_o)$, we get

$$(\mathcal{L})\lim_{n\to\infty} \xi_n = N(0,\ q(\theta_o)/i(\theta_o)). \tag{3.6.34}$$

As ξ_n converges in distribution to $N(0,q(\theta_o)/i(\theta_o))$ and η_n converges in probability to one, we get that ξ_n/η_n converges in distribution to normal with mean zero and variance $q(\theta_o)/i(\theta_o)$ by Lemma 3.2.4. Therefore, from (3.6.25) and (3.6.32), we get that $\beta_n^{1/2}(\hat{\theta}_n - \theta_o)$ converges in distribution to normal with mean zero and variance $\{1+g(\theta_o)/i(\theta_o)\}\{i(\theta_o)\}^{-1}$. Hence the theorem follows.

Definition 3.6.1: An estimator T_n of θ is said to be asymptotically efficient of first-order, if there exist two constants a and b($\neq 0$), not depending on the sample such that

$$(p)\lim_{n\to\infty}\left| \{I_n(\theta)\}^{-1/2} \frac{\partial}{\partial\theta} \log \tilde{p}_n(\theta) - a - b\{I_n(\theta)\}^{-1/2}(T_n-\theta)\right| = 0 \tag{3.6.35}$$

where

$$I_n(\theta) = -E_\theta[\frac{\partial^2}{\partial\theta^2} \log \tilde{p}_n(\theta)].$$

Theorem 3.6.3: Under assumptions (A1)-(A6) a consistent AMLE is asymptotically efficient of first order.

Proof: We have

$$\begin{aligned} -I_n(\theta) &= -E_\theta[\frac{\partial^2}{\partial\theta^2}\log \tilde{p}_n(\theta)] \\ &= -\sum_{i=1}^{\beta_n} E_\theta[\frac{\partial^2}{\partial\theta^2} \log p_{i,n}(\theta)] \end{aligned}$$

by (3.5.4) and (3.5.5). Therefore, by stationarity

$$-I_n(\theta) = \beta_n E_\theta [\frac{\partial^2}{\partial\theta^2} \log p_{1,n}(\theta)].$$

Therefore, by condition (A4),

$$\lim_{n\to\infty} - \beta_1^{-1} I_n(\theta) = i(\theta) > 0. \tag{3.6.36}$$

In view of (3.6.1), (3.6.7), (3.6.36) and by the fact that $\hat{\theta}_n$ is a consistent estimator, the conclusion of the theorem follows as in Theorem 2.4.3.

3.7 Example

In this section, we give an example which satisfies the assumptions of Section 3.5.

Example 3.7.1: Let $\{X_n, n \geq 1\}$ be a sequence of random variables such that (i) $E(X_n) = \mu$ (ii) $Var(X_n) = 1$ and (iii) $Cov(X_n, X_m) = \rho^{|n-m|}$ where $|\rho| < 1$. Suppose that, for all n, the n-dimensional joint density function which depends on a single unknown parameter μ is given by

$$(2\pi)^{-n/2}\{\det \Gamma_n\}^{-1/2} \{\exp[-\{ (\underset{\sim}{x}^{(n)} - \underset{\sim}{\varepsilon}^{(n)})\Gamma_n^{-1}(\underset{\sim}{x}^{(n)} - \underset{\sim}{\varepsilon}^{(n)})^T\}/2] \qquad \ldots(3.7.1)$$

where T denotes the transpose of the vector, $\underset{\sim}{x}^{(n)}$ and $\underset{\sim}{\varepsilon}^{(n)}$ are 1×n vectors $\{x_1, \ldots, x_n\}$ and $\{E(X_1), \ldots, E(X_n)\}$ respectively, and Γ_n^{-1} is the inverse of the variance-covariance matrix of $\underset{\sim}{X}^{(n)}$. The elements of Γ_n are given by (ii) and (iii). Then

$$\Gamma_n^{-1} = (1-\rho^2)^{-1} \begin{bmatrix} 1 & -\rho & 0 & \ldots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \ldots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \ldots & 0 & 0 \\ \vdots & & & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \ldots & -\rho & 1 \end{bmatrix} \qquad (3.7.3)$$

From (3.7.1) and from the variance-covariance matrix it follows that the sequence $\{X_n, n \geq 1\}$ is stationary and ϕ-mixing with

$$\phi_n = \rho^n. \qquad (3.7.3)$$

The problem is to obtain the asymptotic properties of AMLE of μ. We show that the conditions of Section 3.5 are satisfied and thus establish weak consistency, asymptotic normality and first-order efficiency of AMLE $\hat{\mu}_n$ of μ. By Theorem 2.5.1 of Anderson [2], we get that the conditional density function of X_n given $\underset{\sim}{X}^{(n-1)} = \underset{\sim}{x}^{(n-1)}$ to be

$$p_n(\mu) = (2\pi)^{-1/2}\, d_n^{-1} \exp\{-(2d_n^2)^{-1}(x_n-c_n)^2\} \qquad (3.7.4)$$

where

$$c_n = E(X_n) + \Gamma_{1,n-1}\,\Gamma_{n-1}^{-1}\,(\underset{\sim}{x}^{(n-1)} - \underset{\sim}{\varepsilon}^{(n-1)})^T\,, \qquad (3.7.5)$$

$$d_n^2 = \mathrm{Var}(X_n) - \Gamma_{1,n-1}\,\Gamma_{n-1}^{-1}\,\Gamma_{1,n-1}^T \qquad (3.7.6)$$

and $\Gamma_{1,n-1}$ is the $1\times(n-1)$ vector $\{\rho^{n-1}, \rho^{n-2}, \ldots, \rho\}$. We have

$$\begin{aligned}
&\Gamma_{1,n-1}\,\Gamma_{n-1}^{-1}\,(\underset{\sim}{x}^{(n-1)} - \underset{\sim}{\varepsilon}^{(n-1)})^T \\
&= \{\rho^{n-1}, \rho^{n-2}, \ldots, \rho\}\,\Gamma_{n-1}^{-1}\,(\underset{\sim}{x}^{(n-1)} - \underset{\sim}{\varepsilon}^{(n-1)})^T \\
&\doteq \{0, 0, \ldots, \rho\}\,(\underset{\sim}{x}^{(n-1)} - \underset{\sim}{\varepsilon}^{(n-1)})^T \\
&= \rho\,(x_{n-1} + E\,(X_{n-1})).
\end{aligned}$$

Therefore

$$c_n = \mu(1-\rho) + \rho\, x_{n-1} \qquad (3.7.7)$$

and

$$d_n^2 = 1 - \rho^2 . \qquad (3.7.8)$$

Hence (3.7.4) becomes

$$p_n(\mu) = (2\pi)^{-1/2}\,(1-\rho^2)^{-1} \exp[-\{2(1-\rho^2)\}^{-1}(x_n-\mu(1-\rho)-\rho x_{n-1})^2] \qquad \ldots(3.7.9)$$

Clearly,

$$\frac{\partial}{\partial\mu}\log p_n(\mu) = (1+\rho)^{-1}(x_n-\mu(1-\rho)-\rho x_{n-1}) \qquad (3.7.10)$$

and

$$\frac{\partial^2}{\partial\mu^2}\log p_n(\mu) = -(1-\rho)/(1+\rho). \qquad (3.7.11)$$

Conditions (A2), (A3) and (A4) follow from (3.7.10) and (3.7.11). Since $\frac{\partial^2}{\partial\mu^2}\log p_n(\mu)$ is independent of $x_1, x_2, \ldots, x_n$ condition (A5) will be satisfied with $G(x_1, x_2, \ldots, x_n) \equiv 0$. For all choices of the sequences of positive integers $\{\ell_n\}$, condition (A6) follows from (3.7.11). Assumptions (A7) and (A8) follow since

$$\lim_{n\to\infty} E_\mu\left|\frac{\partial}{\partial\mu}\log p_n(\mu)\right|^3 = 2\sqrt{2/\pi}.$$

Since the convergence of ρ^n to zero is always faster than that of $n^{-\theta}$ for all $\theta > 0$ assumption (A8) will be satisfied. Thus all the conditions of Section 3.5 and assumptions (A7) and (A8) are satisfied by this example. Hence, by Theorems 3.6.1, 3.6.2 and 3.6.3, it follows that the AMLE $\hat{\mu}_n$ of μ is consistent, asymptotically normal and first-order efficient.

3.8 Conclusions.

In this chapter, we proved weak consistency, asymptotic normality and first-order efficiency of an AMLE for stationary ϕ-mixing processes with some assumptions on the mixing coefficient ϕ. It is easy to see from Definition 3.2.1 that, if $\phi_n \equiv 0$ for all n, then

ϕ-mixing condition reduces to independence. The assumptions of Section 3.5 will be satisfied if we choose the sequence of non-negative integers $\{\ell_n\}$ to be identically zero and AMLE in this case reduces to the usual MLE. If $\phi_n = 0$ for all $n > m$, then the sequence of random variables $\{X_n, n \geq 1\}$ are m-dependent. Thus m-dependence is a particular case of ϕ-mixing process. Hence if a sequence of stationary m-dependent random variables satisfies the conditions of Section 3.5, then the conclusions of Theorem 3.6.1 to 3.6.3 hold in this case also. The example given in Section 2.5 satisfies all the assumptions of Section 3.5 except assumption (A6). For, in this case there exists no sequence $\{\ell_n\}$ of non-negative integers such that $\lim_{n\to\infty} \ell_n = c$ finite and

$$\lim_{n\to\infty} E_\theta[\{\frac{\partial}{\partial\theta} \log p(x_{1+\ell_n} | x_1, \ldots, x_{\ell_n})\}^2] = i(\theta).$$

CHAPTER IV

MAXIMUM LIKELIHOOD ESTIMATION FOR DEPENDENT RANDOM VARIABLES

4.1 Introduction :

In the previous two chapters, we studied some large sample properties of maximum likelihood estimators when the observations are from a sequence of stationary random variables which are asymptotically independent in some sense. In this chapter, we obtain strong consistency, asymptotic normality and first-order efficiency of an MLE for arbitrary stochastic processes satisfying some regularity conditions. The results in Loéve [28] regarding the strong law of large numbers and central limit theorem for dependent random variables are used to obtain strong consistency and asymptotic normality of MLE respectively. Bhat and Kulkarni [5] applied similar result of Loéve [28] for double sequences in a different problem.

We state some preliminary results in Section 4.2. The regularity conditions to be satisfied by the n-dimensional joint density function are given in Section 4.3. As in Bar-Shalom [4], these conditions are expressed in terms of the probability density of observations conditioned upon all past observations. The proofs of strong consistency, asymptotic normality and first-order efficiency of MLE are given in Section 4.4. As an application of these results, we obtain the Bernstein-vonMises theorem for dependent random variables, in Section 4.5. Section 4.6 contains some examples which satisfy the regularity conditions presented in Section 4.3.

4.2 Preliminary Results

In this section, we state some lemmas which will be used later in the chapter.

Lemma 4.2.1: If the random variables X_k, $k = 1,2,\ldots,n$ are such that $\sum_{n=1}^{\infty} b_n^{-2} \operatorname{Var} X_n < \infty$ with $\lim_{n\to\infty} b_n = \infty$, then

$$(a.s)\lim_{n\to\infty} b_n^{-1} \sum_{k=1}^{n} \{X_k - E(X_k|X_1,\ldots,X_{k-1})\} = 0. \qquad (4.2.1)$$

Proof: For proof, we refer to Loéve [28], p. 387.

Lemma 4.2.2: Let $X_{n1}, X_{n2}, \ldots$ be random variables centered at their expectations. Under the Liapunov's condition

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} E|X_{nk}|^{2+\delta} = 0, \qquad (4.2.2)$$

if

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} E|E' X_{nk}| = 0 \qquad (4.2.3)$$

and

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} E|E' X_{nk}^2 - E X_{nk}^2| = 0, \qquad (4.2.4)$$

then

$$(\mathcal{L})\lim_{n\to\infty} (s_n^{-1} \sum_{k=1}^{k_n} X_{nk}) = N(0, 1) \tag{4.2.5}$$

where

$$s_n^2 = E(\sum_{k=1}^{k_n} X_{nk})^2$$

and N(0, 1) is the normal distribution with mean zero and variance one and E' denotes the conditional expectation with respect to $X_{no} + \ldots + X_{n,k_n-1}$.

Proof: For proof, we refer to Loéve [28] p. 377.

Lemma 4.2.3: Let $\{X_k, k \geq 1\}$ be a sequence of random variables satisfying the following conditions.

(i) $\lim_{n\to\infty} n^{-(1+\delta/2)} \sum_{k=1}^{n} E|X_k|^{2+\delta} = 0$,

for some $\delta > 0$,

(ii) $E(X_k | X_1, \ldots, X_{k-1}) = 0$

almost surely, and

(iii) $E(X_k^2 | X_1, \ldots, X_{k-1}) = \sigma_k^2$

almost surely, where σ_k^2 is independent of $X_1, \ldots, X_{k-1}$.

Then,

$$(\mathcal{L})\lim_{n\to\infty} n^{-1/2} s_n^{-1} \sum_{k=1}^{n} X_k = N(0, 1)$$

where

$$s_n^2 = n^{-1} \sum_{k=1}^{n} E\, X_k^2 = n^{-1} \sum_{k=1}^{n} \sigma_k^2 \tag{4.2.6}$$

and N(0, 1) is the standard normal distribution.

Proof: Since the σ-field generated by the random variable $X_1 + \ldots + X_{k-1}$ is contained in the σ-field generated by the random variables $X_1, \ldots, X_{k-1}$, the assumptions (4.2.2) to (4.2.5) of Lemma 4.2.2 will be satisfied if we take $X_{nk} = n^{-1/2} X_k$ for all n and $1 \leq k \leq n$. Thus, by an application of Lemma 4.2.2, the required result follows.

Lemma 4.2.4: If g is a continuous function mapping s-dimensional Euclidean space into itself with the property that, for every $\underset{\sim}{\theta}$ such that $\|\underset{\sim}{\theta}\| = 1$, $\underset{\sim}{\theta}' \, g(\underset{\sim}{\theta}) < 0$, then there exists a point $\hat{\underset{\sim}{\theta}}$ such that $\|\hat{\underset{\sim}{\theta}}\| < 1$ and $\underset{\sim}{g}(\hat{\underset{\sim}{\theta}}) = 0$, where $\underset{\sim}{\theta}' \, \underset{\sim}{g}(\underset{\sim}{\theta}) = \sum_{t=1}^{s} \theta_t \, g_t(\underset{\sim}{\theta})$.

Proof: For proof, we refer to Aitchison and Silvey [1], Lemma 2.

Lemma 4.2.5: Suppose $\{U_n, n \geq 1\}$ is a sequence of random variables such that $U_n \to U$ in distribution. Further suppose that $\{V_n, n \geq 1\}$ is another sequence of random variables such that

$$|U_n - V_n| \leq \varepsilon_n \, |V_n|$$

where ε_n tends to zero in probability as n tends to infinity. Then V_n converges in distribution to U.

Proof: For proof, see Billingsley [8], p. 62.

Lemma 4.2.6: Let $X_1, \ldots, X_k$ be k random variables. Then for all $r \geq 1$,

$$E\left|\sum_{i=1}^{k} X_i\right|^r \leq 2^{k(r-1)} \sum_{i=1}^{k} E|X_i|^r.$$

Proof: The lemma follows by repeated application of C_r-inequality of Loéve [28].

4.3 Regularity Conditions:

Let the set of dependent observations be

$$y^n = \{y_1, \ldots, y_n\} \tag{4.3.1}$$

where y_k are real valued random variables, with a known joint probability density function with respect to a σ-finite product measure μ^n. This density function depends on a ℓ-dimensional unknown parameter $\underset{\sim}{\theta}$, that is

$$p(y_1, \ldots, y_n; \underset{\sim}{\theta}) \equiv p(y^n|\underset{\sim}{\theta}). \tag{4.3.2}$$

The true value $\underset{\sim}{\theta}^0$ of the parameter $\underset{\sim}{\theta}$ is estimated by a Borel-measurable function $\hat{\underset{\sim}{\theta}}_n(y^n)$ obtained by maximizing the likelihood function

$$p(y^n|\underset{\sim}{\theta}) = \prod_{k=1}^{n} p_k(\underset{\sim}{\theta}) \tag{4.3.3}$$

where

$$p_k(\underset{\sim}{\theta}) = p(y_k|y^{k-1}, \underset{\sim}{\theta}). \tag{4.3.4}$$

We denote $\log p_k(\underset{\sim}{\theta})$ by $g(k;\underset{\sim}{\theta})$, the first partial derivative of $g(k;\underset{\sim}{\theta})$ with respect to θ_u by $g_u(k;\underset{\sim}{\theta})$ and the partial derivative of $g_u(k;\underset{\sim}{\theta})$ with respect to θ_v by $g_{uv}(k;\underset{\sim}{\theta})$. Suppose the following regularity conditions are satisfied for all k.

(E1) The parameter space Θ is an open subset of the ℓ-dimensional Euclidean space E^{ℓ}.

(E2) For almost all y^{k-1} and for all $\underset{\sim}{\theta}\varepsilon\Theta$, $g_u(k;\underset{\sim}{\theta})$ and $g_{uv}(k;\underset{\sim}{\theta})$ exists for $u,v = 1,\ldots,\ell$ and are continuous throughout Θ.

(E3) For almost all y^{k-1} and for all $\underset{\sim}{\theta}\ \varepsilon\ \Theta$,

$$E_{\underset{\sim}{\theta}}\ [g_u(k;\underset{\sim}{\theta})\,|y^{k-1}] = 0, \quad u=1,\ldots,\ell \tag{4.3.5}$$

and

$$- E_{\underset{\sim}{\theta}}[g_{uv}(k;\underset{\sim}{\theta})\,|y^{k-1}] = E_{\underset{\sim}{\theta}}[g_u(k;\underset{\sim}{\theta})\ g_v(k;\underset{\sim}{\theta})\,|y^{k-1}]$$

$$= \sigma^{k}_{uv}(\underset{\sim}{\theta}) \tag{4.3.6}$$

for $u,v = 1,\ldots,\ell$, where $\sigma^{k}_{uv}(\underset{\sim}{\theta})$ is independent of $y_1,\ldots,y_{k-1}$.

(E4) For all $\underset{\sim}{\theta}\ \varepsilon\ \Theta$

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \sigma^{k}_{uv}(\underset{\sim}{\theta}) = \sigma_{uv}(\underset{\sim}{\theta}) < \infty$$

and the $\ell\times\ell$ matrix

$$\Sigma(\underset{\sim}{\theta}) = ((\sigma_{uv}(\underset{\sim}{\theta})\ )) \tag{4.3.7}$$

is non-singular.

(E5) For all $\underset{\sim}{\theta}\ \varepsilon\ \Theta$, there exists a neighbourhood $V(\underset{\sim}{\theta})$ such that for all $\underset{\sim}{\theta}'\ \varepsilon\ V(\underset{\sim}{\theta})$

$$|g_{uv}(\underset{\sim}{\theta}') - g_{uv}(\underset{\sim}{\theta})| \le |\underset{\sim}{\theta}'-\underset{\sim}{\theta}|_{\ell}\ G(y^k, \underset{\sim}{\theta}) \tag{4.3.8}$$

for all $u,v = 1,\ldots,\ell$, where $G(y^k,\underset{\sim}{\theta}) \geq 0$,

$$(\text{a.s})\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} E_{\underset{\sim}{\theta}}[G(y^k;\underset{\sim}{\theta})\,|y^{k-1}] = M(\underset{\sim}{\theta}) < \infty,$$

$M(\underset{\sim}{\theta})$ is a constant and $|\cdot|_\ell$ denotes the usual Euclidean norm in E^ℓ.

(E6) For all $\underset{\sim}{\theta}\varepsilon\Theta$ and for all $u,v = 1, \ldots, \ell$

$$\text{(i)} \quad \sum_{k=1}^{\infty} k^{-2}\, \text{Var}_{\underset{\sim}{\theta}}[g_u(k;\underset{\sim}{\theta})] < \infty, \qquad (4.3.9)$$

$$\text{(ii)} \quad \sum_{k=1}^{\infty} k^{-2}\, \text{Var}_{\underset{\sim}{\theta}}[g_{uv}(k;\underset{\sim}{\theta})] < \infty, \qquad (4.3.10)$$

and

$$\text{(iii)} \quad \sum_{k=1}^{\infty} k^{-2}\, \text{Var}_{\underset{\sim}{\theta}}[G(y^k;\underset{\sim}{\theta})] < \infty. \qquad (4.3.11)$$

The following condition is weaker than (E6).

(E6') For all $\underset{\sim}{\theta} \varepsilon \Theta$, and for all $u,v = 1,\ldots,\ell$

$$\text{(i)} \quad \lim_{k\to\infty} k^{-1}\, \text{Var}_{\underset{\sim}{\theta}}[g_u(k;\underset{\sim}{\theta})] < \infty, \qquad (4.3.12)$$

$$\text{(ii)} \quad \lim_{k\to\infty} k^{-1}\, \text{Var}_{\underset{\sim}{\theta}}[g_{uv}(k;\underset{\sim}{\theta})] < \infty, \qquad (4.3.13)$$

and

$$\text{(iii)} \quad \lim_{k\to\infty} k^{-1}\, \text{Var}_{\underset{\sim}{\theta}}[G(y^k;\underset{\sim}{\theta})] < \infty. \qquad (4.3.14)$$

(E7) For all $\underset{\sim}{\theta} \varepsilon \Theta$ and for all $u=1,\ldots,\ell$

$$\lim_{n\to\infty} n^{-(1+\delta/2)} \sum_{k=1}^{n} E_{\underset{\sim}{\theta}}\, |\, g_u(k;\underset{\sim}{\theta})|^{2+\delta} = 0 \qquad (4.3.15)$$

for some $\delta > 0$, independent of $\underset{\sim}{\theta}$.

The maximum likelihood estimate of $\underset{\sim}{\theta}$ is defined as a solution of the system of likelihood equations

$$\frac{\partial}{\partial\theta_u} \log p(y^n|\underset{\sim}{\theta}) = 0, \; u = 1,\ldots,\ell, \tag{4.3.16}$$

in a neighbourhood of the true parameter value $\underset{\sim}{\theta}^0$.

4.4 The Asymptotic Properties of MLE:

Theorem 4.4.1: If assumptions (E1) - (E6) of Section 4.3 are satisfied, then the system of likelihood equations (4.3.16) has a root, which is strongly consistent.

Proof: For all $t = 1,\ldots,\ell$, expanding $\frac{\partial}{\partial\theta_t} \log p(y^n|\underset{\sim}{\theta})$ about the true value $\underset{\sim}{\theta}^0$ of the parameter $\underset{\sim}{\theta}$, it follows that

$$\frac{\partial}{\partial\theta_t} \log p(y^n|\underset{\sim}{\theta}) = \frac{\partial}{\partial\theta_t} \log p(y^n|\underset{\sim}{\theta}^0) +$$

$$+ \sum_{s=1}^{\ell} (\theta_s - \theta_s^0) \frac{\partial^2}{\partial\theta_t \partial\theta_s} \log p(y^n|\underset{\sim}{\theta}') \qquad \ldots(4.4.1)$$

where $\theta_s' = \theta_s^0 + \alpha_s(\theta_s - \theta_s^0)$, $|\alpha_s| < 1$, $s=1,\ldots,\ell$.

The right side of (4.4.1) can be written as

$$\sum_{k=1}^{n} g_t(k,\underset{\sim}{\theta}^0) + \sum_{k=1}^{n} \sum_{s=1}^{\ell} (\theta_s - \theta_s^0)\, g_{ts}(k,\underset{\sim}{\theta}^0)$$

$$+ \sum_{k=1}^{n} \sum_{s=1}^{\ell} (\theta_s - \theta_s^0)\{g_{ts}(k,\underset{\sim}{\theta}') - g_{ts}(k,\underset{\sim}{\theta}^0)\}.$$

It follows from (E5) and (E6) there exists a neighbourhood $V(\underset{\sim}{\theta}^0)$ such that for $\underset{\sim}{\theta}' \in V(\underset{\sim}{\theta}^0)$,

$$|g_{ts}(k,\underset{\sim}{\theta}') - g_{ts}(k,\underset{\sim}{\theta}^0)| \leq |\underset{\sim}{\theta}' - \underset{\sim}{\theta}^0|_{\ell}\, G(y^k;\underset{\sim}{\theta}^0)$$

where $G(y^k;\underset{\sim}{\theta}^0) \geq 0$ and

$$(\mathrm{a.s})\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} E_{\underset{\sim}{\theta}^0} [G(y^k;\underset{\sim}{\theta}^0) \,|y^{k-1}] = M(\underset{\sim}{\theta}^0)=M_o \text{ (say).}$$

Clearly,

$$\Big|\sum_{k=1}^{n} \sum_{s=1}^{\ell} (\theta_s-\theta_s^0)\{g_{ts}(k,\underset{\sim}{\theta}') - g_{ts}(k,\underset{\sim}{\theta}^0)\}\Big|$$

$$\leq \sum_{k=1}^{n} \sum_{s=1}^{\ell} |\theta_s-\theta_s^0|\,|\underset{\sim}{\theta}-\underset{\sim}{\theta}^0|_\ell\, G(y_k;\underset{\sim}{\theta}^0)$$

$$\leq \ell\, |\underset{\sim}{\theta}-\underset{\sim}{\theta}^0|_\ell^2 \sum_{k=1}^{n} G(y^k,\underset{\sim}{\theta}^0)$$

$$\leq \ell^2\, |\underset{\sim}{\theta}-\underset{\sim}{\theta}^0|_\ell^2 \sum_{k=1}^{n} G(y^k;\underset{\sim}{\theta}^0)\ . \tag{4.4.2}$$

Hence we get

$$n^{-1} \frac{\partial}{\partial\theta_t} \log p(y^n|\underset{\sim}{\theta}) = n^{-1} \sum_{k=1}^{n} g_t(k,\underset{\sim}{\theta}^0) + \sum_{s=1}^{\ell} (\theta_s-\theta_s^0)\{n^{-1} \sum_{k=1}^{n} g_{ts}(k,\underset{\sim}{\theta}^0)\} + \beta_n\, \ell^2\, |\underset{\sim}{\theta}-\underset{\sim}{\theta}^0|_\ell^2\, n^{-1} \sum_{k=1}^{n} G(y^k;\underset{\sim}{\theta}^0) \tag{4.4.3}$$

where $|\beta_n| \leq 1$ for all n. By (E3) and (E4), it is clear that for any $t = 1,\ldots,\ell$,

$$(\mathrm{a.s})\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} E_{\underset{\sim}{\theta}^0}\{g_t(k,\underset{\sim}{\theta}^0)|y^{k-1}\} = 0, \tag{4.4.4}$$

$$(\mathrm{a.s})\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} E_{\underset{\sim}{\theta}^0} \{g_{ts}(k,\underset{\sim}{\theta}^0)|y^{k-1}\} = -\,\sigma_{ts}(\underset{\sim}{\theta}^0)\ , \tag{4.4.5}$$

for $t, s=1,\ldots,\ell$, and

$$\text{(a.s)} \lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} E_{\underset{\sim}{\theta}^0}\{G(y^k;\underset{\sim}{\theta}^0)\,|\,y^{k-1}\} = M_o. \qquad (4.4.6)$$

Therefore, by (E6) and Lemma 4.2.1, for $t,s = 1,\ldots,\ell$,

$$\text{(a.s)} \lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} g_t(k,\underset{\sim}{\theta}^0) = 0\ , \qquad (4.4.7)$$

$$\text{(a.s)} \lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} g_{ts}(k,\underset{\sim}{\theta}^0) = -\sigma_{ts}\ , \qquad (4.4.8)$$

and

$$\text{(a.s)} \lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} G(y^k;\underset{\sim}{\theta}^0) = M_o\ , \qquad (4.4.9)$$

where $\sigma_{ts} = \sigma_{ts}(\underset{\sim}{\theta}^0)$. The $\ell\times\ell$ matrix $\Sigma_o = \Sigma(\underset{\sim}{\theta}^0)$ is non-singular by condition (E4) and hence positive definite. It follows that there is a $\tau > 0$ such that $\sum_{t=1}^{\ell} \sum_{s=1}^{\ell} \sigma_{ts}\, u_t\, u_s \geq \tau$ for any vector u in E^ℓ such that $|u|_\ell = 1$. Suppose $\varepsilon > 0$ is given. Choose $\delta = \delta(\varepsilon) > 0$ such that

$$\delta < \varepsilon, \{\underset{\sim}{\theta} : |\underset{\sim}{\theta}-\underset{\sim}{\theta}^0|_\ell \leq \delta\} \subset V(\underset{\sim}{\theta}^0) \text{ and } \delta < \tau\{3\ell^3(M+1)\}^{-1}. \qquad \ldots (4.4.10)$$

Let $S^{(t)}$, $S_s^{(t)}$ and S_3, $t,s = 1,\ldots,\ell$ be sets such that

$$S^{(t)} = \bigcup_{n=n_o}^{\infty} \{\underset{\sim}{y} : |n^{-1} \sum_{k=1}^{n} g_t(k,\underset{\sim}{\theta}^0)| > \delta^2\},$$

$$S_s^{(t)} = \bigcup_{n=n_o}^{\infty} \{\underset{\sim}{y} : |n^{-1} \sum_{k=1}^{n} g_{ts}(k,\underset{\sim}{\theta}^0)+\sigma_{ts}| > \delta\}\ ,$$

and

$$S_3 = \bigcup_{n=n_o}^{\infty} \{\underset{\sim}{y} : n^{-1} \sum_{k=1}^{n} G(y^k;\underset{\sim}{\theta}^0) > M_o+1\}\ .$$

By (4.4.7) and (4.4.9) and by the definition of almost sure convergence, there exist positive numbers $n_1(\varepsilon,t)$, $n_2(\varepsilon,t,s)$ and $n_3(\varepsilon)$ such that, for all $t,s = 1,\ldots,\ell$, $P(S^{(t)}) \leq \varepsilon/(3\ell)$ for $n_o > n_1(\varepsilon,t)$, $P(S_s^{(t)}) \leq \varepsilon/(3\ell^2)$ for $n_o > n_2(\varepsilon,t,s)$ and $P(S_3) \leq \varepsilon/3$ for $n_o > n_3(\varepsilon)$. Then, for

$$n_o > N(\varepsilon) = \max\{ n_1(\varepsilon,t);\ n_2(\varepsilon,t,s);\ n_3(\varepsilon), t,s=1,\ldots,\ell\}$$

all the above $(\ell+\ell^2+1)$ inequalities will be satisfied. By (4.4.3), we have

$$\begin{aligned}
&\Big|n^{-1} \sum_{k=1}^{n} g_t(k,\underset{\sim}{\theta}) + \sum_{s=1}^{\ell} \sigma_{ts} (\theta_s - \theta_s^0)\Big| \\
&\quad = \Big|n^{-1} \sum_{k=1}^{n} g_t(k,\underset{\sim}{\theta}^0) + \sum_{s=1}^{\ell} (\theta_s - \theta_s^0) \{n^{-1} \sum_{k=1}^{n} g_{ts}(k,\underset{\sim}{\theta}^0) + \sigma_{ts}\} \\
&\qquad + \beta_n\, \ell^2 |\underset{\sim}{\theta} - \underset{\sim}{\theta}^0|_\ell^2\, n^{-1} G(y^k;\underset{\sim}{\theta}^0)\Big| \\
&\quad \leq \Big|n^{-1} \sum_{k=1}^{n} g_t(k,\underset{\sim}{\theta}^0)\Big| + \sum_{s=1}^{\ell} |\theta_s - \theta_s^0|\, \Big|n^{-1} \sum_{k=1}^{n} g_{ts}(k,\underset{\sim}{\theta}^0) + \sigma_{ts}\Big| \\
&\qquad + \ell^2 |\underset{\sim}{\theta} - \underset{\sim}{\theta}^0|_\ell^2\, n^{-1} G(y^k;\underset{\sim}{\theta}^0). \qquad (4.4.11)
\end{aligned}$$

Let S = union of all the $(\ell+\ell^2+1)$ sets $S^{(t)}$, $t=1,\ldots,\ell$, $S_s^{(t)}$, $t,s = 1,\ldots,\ell$ and S_3. Then on $\bar{S}$, if $|\underset{\sim}{\theta} - \underset{\sim}{\theta}^0|_\ell \leq \delta$, for all $t = 1,\ldots,\ell$, the right side of (4.4.11) is

$$\begin{aligned}
&\leq \delta^2 + \delta \sum_{s=1}^{\ell} |\theta_s - \theta_s^0| + \ell^2 |\underset{\sim}{\theta} - \underset{\sim}{\theta}^0|_\ell^2 (M_o+1) \\
&\leq \delta^2 + \delta\ell |\underset{\sim}{\theta} - \underset{\sim}{\theta}^0|_\ell + \ell^2 |\underset{\sim}{\theta} - \underset{\sim}{\theta}^0|_\ell^2 (M_o+1) \\
&\leq 3\ell^2 (M_o+1)\, \delta^2. \qquad (4.4.12)
\end{aligned}$$

Hence, by (4.4.10), if $|\underset{\sim}{\theta}-\underset{\sim}{\theta}^{0}|_{\ell} = \delta$, then

$$\sum_{t=1}^{\ell} (\theta_t-\theta_t^0)[n^{-1}\sum_{k=1}^{n} g_t(k,\underset{\sim}{\theta}^0)]$$

$$\leq - \sum_{t=1}^{\ell}\sum_{s=1}^{\ell} \sigma_{ts} (\theta_t-\theta_t^0)(\theta_s-\theta_s^0) + 3\ell^3 (M_o+1)\delta^3$$

$$\leq - |\underset{\sim}{\theta}-\underset{\sim}{\theta}^{0}|_{\ell}^{2} \sum_{t=1}^{\ell}\sum_{s=1}^{\ell} \sigma_{ts} \frac{(\theta_t-\theta_t^0)}{|\underset{\sim}{\theta}-\underset{\sim}{\theta}^{0}|_{\ell}} \frac{(\theta_s-\theta_s^0)}{|\underset{\sim}{\theta}-\underset{\sim}{\theta}^{0}|_{\ell}} + 3\ell^3(M_o+1)\delta^3$$

$$\leq - \delta^2\tau + 3\ell^3(M_o+1)\delta^3 < 0 \qquad (4.4.13)$$

since $\delta < \tau\{3\ell^3(M_o+1)\}^{-1}$. It follows from Lemma 4.2.4 that there is a value $\hat{\underset{\sim}{\theta}}_n$ of $\underset{\sim}{\theta}$ such that

$$n^{-1}\frac{\partial}{\partial\theta_t}\log p(y^n|\hat{\underset{\sim}{\theta}}_n) = n^{-1}\sum_{k=1}^{n} g_t(k, \hat{\underset{\sim}{\theta}}_n) = 0 \qquad (4.4.14)$$

for $t = 1,\ldots,\ell$, and such that $|\hat{\underset{\sim}{\theta}}_n-\underset{\sim}{\theta}^0|_{\ell} \leq \delta < \varepsilon$. Since,

$$\bar{S} \subset \bigcap_{n=n_o}^{\infty} \{\underset{\sim}{y} : \sum_{t=1}^{\ell}(\theta_t-\theta_t^0)[n^{-1}\sum_{k=1}^{n} g_t(k,\underset{\sim}{\theta}^0)] < 0\}$$

for $n > N(\varepsilon)$ and for δ as in (4.4.10), it follows that, for $n > N(\varepsilon)$ and for δ as in (4.4.10),

$$P[\bigcup_{n=n_o}^{\infty} \{\underset{\sim}{y} : \sum_{t=1}^{\ell}(\theta_t-\theta_t^0)[n^{-1}\sum_{k=1}^{n} g_t(k;\underset{\sim}{\theta}^0)] \geq 0\}] = P[S]$$

$$\leq \sum_{t=1}^{\ell} P(S^{(t)}) + \sum_{t=1}^{\ell}\sum_{s=1}^{\ell} P(S_s^{(t)}) + P(S_3)$$

$$\leq \ell\varepsilon/(3\ell) + \ell^2\varepsilon/(3\ell) + \varepsilon/3 = \varepsilon .$$

Hence, by (4.4.13) and Lemma 4.2.4 we get that, if δ is as in (4.4.10), then there is a value $\hat{\underset{\sim}{\theta}}_n$ of $\underset{\sim}{\theta}$ such that

$$\frac{\partial}{\partial\theta_t} \log p(y^n|\hat{\underset{\sim}{\theta}}_n) = 0$$

almost surely for $t = 1,\ldots,\ell$ and such that $|\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}^0|_\ell < \delta < \varepsilon$ for large n. This proves the existence of a strongly consistent solution of the system of likelihood equations (4.3.16).

Remark 4.4.1: If the assumptions (E1) - (E5) and (E6') instead of (E6) are satisfied, then by an argument similar to that given by Billingsley [8] in Markov case, it can be shown that, the MLE is weakly consistent.

Theorem 4.4.2: Under the assumptions (E1) - (E5), (E6') and (E7), any weakly consistent root of the system of likelihood equations (4.3.16), is asymptotically normal.

Proof: Let $\underset{\sim}{y}(n) = \{y_1(n), y_2(n), \ldots, y_\ell(n)\}$ be the random vector with components

$$y_t(n) = n^{-1/2} \sum_{k=1}^{n} g_t(k, \underset{\sim}{\theta}^0). \tag{4.4.15}$$

If $\hat{\underset{\sim}{\theta}}_n$ is a weakly consistent solution of the likelihood equations, let $\underset{\sim}{u}(n) = \{u_1(n),\ldots,u_\ell(n)\}$ be the random vector with components

$$u_t(n) = n^{1/2} (\hat{\theta}_{tn} - \theta_t^0). \tag{4.4.16}$$

To prove that the random vector $\underset{\sim}{y}(n)$ converges in distribution to normal with mean vector zero and variance-covariance matrix $\Sigma_0 = \Sigma(\underset{\sim}{\theta}^0)$, by standard Cramér-Wold result [15], it is enough to show that for any set $q_1, \ldots, q_\ell$ of real numbers

$$(\mathcal{L}) \lim_{n\to\infty} \left[\sum_{t=1}^{\ell} q_t \, y_t(n) \right] = N(0, \beta^2) \tag{4.4.17}$$

where

$$\beta^2 = \sum_{t=1}^{\ell} \sum_{s=1}^{\ell} \sigma_{ts} \, q_t \, q_s . \tag{4.4.18}$$

We have

$$z_n = \sum_{t=1}^{\ell} q_t \, y_t(n) = n^{-1/2} \sum_{k=1}^{n} \sum_{t=1}^{\ell} q_t \, g_t(k, \underset{\sim}{\theta}^0)$$

$$= n^{-1/2} \sum_{k=1}^{n} u_k \tag{4.4.19}$$

where

$$u_k = \sum_{t=1}^{\ell} q_t \, g_t(k; \underset{\sim}{\theta}^0) \quad . \tag{4.4.20}$$

Let, $\mathcal{F}_k$ be the σ-field generated by $y_1, \ldots, y_k$, for $k \geq 1$. Then by (E3) and (E7)

$$E_{\underset{\sim}{\theta}^0} [u_k | \mathcal{F}_{k-1}] = 0, \tag{4.4.21}$$

$$E_{\underset{\sim}{\theta}^0} [u_k^2 | \mathcal{F}_{k-1}] = \sum_{t=1}^{\ell} \sum_{s=1}^{\ell} q_t \, q_s \, \sigma_{ts}^k , \tag{4.4.22}$$

where σ_{ts}^k is independent of $y_1, \ldots, y_{k-1}$, and

$$\lim_{n\to\infty} n^{-(1+\delta/2)} \sum_{k=1}^{n} E_{\underset{\sim}{\theta}^0} (|u_k|^{2+\delta})$$

$$= \lim_{n\to\infty} n^{-(1+\delta/2)} \sum_{k=1}^{n} E_{\underset{\sim}{\theta}^0} \left| \sum_{t=1}^{\ell} q_t \, g_t(k; \underset{\sim}{\theta}^0) \right|^{2+\delta}$$

$$\leq \lim_{n\to\infty} n^{-(1+\delta/2)} \, 2^{\ell(1+\delta)} \sum_{k=1}^{n} \sum_{t=1}^{\ell} |q_t|^{2+\delta} E_{\underset{\sim}{\theta}^0} |g_t(k; \underset{\sim}{\theta}^0)|^{2+\delta} .$$

The last step follows from Lemma 4.2.6. By assumption (E7) it follows that,

$$\lim_{n\to\infty} n^{-(1+\delta/2)} \sum_{k=1}^{n} E_{\underset{\sim}{\theta}^0}(|u_k|^{2+\delta}) = 0. \tag{4.4.23}$$

It is clear from (4.4.21) - (4.4.23) that the conditions of Lemma 4.2.3 are satisfied and hence

$$(\mathcal{L}) \lim_{n\to\infty} [n^{-1/2} s_n^{-1} \sum_{k=1}^{n} u_k] = N(0,1) \tag{4.4.24}$$

where

$$s_n^2 = \frac{1}{n} \sum_{k=1}^{n} E(u_k^2)$$

$$= \frac{1}{n} \sum_{k=1}^{n} [\sum_{t=1}^{\ell} \sum_{s=1}^{\ell} q_t\, q_s\, \sigma_{ts}^k]$$

$$= \sum_{t=1}^{\ell} \sum_{s=1}^{\ell} q_t\, q_s \{n^{-1} \sum_{k=1}^{n} \sigma_{ts}^k\}. \tag{4.4.25}$$

Since s_n^2 tends to β^2 in (4.4.18) as $n \to \infty$ by condition(E4) we get from (4.4.24) that

$$(\mathcal{L}) \lim_{n\to\infty} n^{-1/2} \sum_{k=1}^{n} u_k = N(0,\beta^2). \tag{4.4.26}$$

Therefore by Cramér-Wold theorem [15], it follows that the random vector $\underset{\sim}{y}(n)$ converges in distribution to normal with mean vector zero and variance-covariance matrix Σ_o that is,

$$(\mathcal{L}) \lim_{n\to\infty} \underset{\sim}{y}(n) = N(\underset{\sim}{0}, \Sigma_o). \tag{4.4.27}$$

Let the neighbourhood $V(\underset{\sim}{\theta}^0)$ and the function G be as in the preceeding theorem. If $\hat{\underset{\sim}{\theta}}_n$ is a consistent solution of the likelihood equation, then $\hat{\underset{\sim}{\theta}}_n$ lies

$\{\hat{\underset{\sim}{\theta}} : |\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}^0|_\ell \leq \delta\}$ almost surely. Hence by (4.4.3), (4.4.15) and (4.4.16) we have that

$$y_t(n) + \sum_{s=1}^{\ell} u_s(n) \, [n^{-1} \sum_{k=1}^{n} g_{ts}(k, \underset{\sim}{\theta}^0)]$$

$$+ \beta_n \ell^2 |\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}^0|_\ell \cdot |\underset{\sim}{u}(n)|_\ell \cdot [n^{-1} \sum_{k=1}^{n} G(y^k; \underset{\sim}{\theta}^0)] = 0 \qquad (4.4.28)$$

for $t = 1, \ldots, \ell$, where $|\beta_n| < 1$. From (4.4.4) - (4.4.6) and the fact that $|\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}^0|_\ell \to 0$ a.s as $n \to \infty$, it follows that

$$y_t(n) - \sum_{s=1}^{\ell} u_s(n) \, (\sigma_{ts} + \varepsilon_{ts}(n))$$

$$= -\beta_n \ell^2 \, \gamma_n \, |\underset{\sim}{u}(n)|_\ell \, (M_0 + \gamma_n') \qquad (4.4.29)$$

where $\varepsilon_{ts}(n)$, γ_n and γ_n' are sequences which tend to zero almost surely as $n \to \infty$. Therefore,

$$|\underset{\sim}{y}(n) - \Sigma_0 \, \underset{\sim}{u}(n)|_\ell \leq \varepsilon_n \, |\underset{\sim}{u}(n)|_\ell \qquad (4.4.30)$$

where $\varepsilon_n \to 0$ almost surely as $n \to \infty$. Hence, by Lemma 4.2.5, the distributions of $\underset{\sim}{y}(n)$ and $\Sigma_0 \, \underset{\sim}{u}(n)$ are asymptotically the same. Therefore, by (4.4.27), it follows that

$$(\mathcal{L}) \lim_{n \to \infty} \{\Sigma_0 \, \underset{\sim}{u}(n)\} = N(\underset{\sim}{0}, \Sigma_0)$$

so that

$$(\mathcal{L}) \lim_{n \to \infty} \{\underset{\sim}{u}(n)\} = N(0, \Sigma_0^{-1}) \qquad (4.4.31)$$

Hence the theorem follows from (4.4.16).

As an application of the above theorem, we prove the following result:

Theorem 4.4.3: Under assumptions (E1) - (E7),

$$(\mathcal{L})\lim_{n\to\infty} [2\{\max_{\underset{\sim}{\theta}\in\Theta} \ln p(y^n|\underset{\sim}{\theta}) - \ln p(y^n|\underset{\sim}{\theta}^0)\}] = \chi^2_\ell$$

where χ^2_ℓ is the central Chi-square distribution with ℓ degrees of freedom.

Proof: By the mean value theorem, we have

$$\begin{aligned} g(k;\underset{\sim}{\theta}) = g(k;\underset{\sim}{\theta}^0) &+ \sum_{t=1}^{\ell} (\theta_t-\theta_t^0)\, g_t(k;\underset{\sim}{\theta}^0) \\ &+ \frac{1}{2}\sum_{t=1}^{\ell}\sum_{s=1}^{\ell} (\theta_t-\theta_t^0)\,(\theta_s-\theta_s^0)\, g_{ts}(k;\underset{\sim}{\theta}^0) \\ &+ \beta_n \ell^3 |\underset{\sim}{\theta}-\underset{\sim}{\theta}^0|_\ell^3 G(y^k;\underset{\sim}{\theta}^0) \qquad (4.4.32) . \end{aligned}$$

where $G(y^k;\underset{\sim}{\theta}^0)$ is as in Theorem 4.4.1 and $|\beta_n| < 1/6$. Since $\hat{\underset{\sim}{\theta}}$ is a consistent root of the likelihood equation, (4.4.32) implies that

$$\begin{aligned} 2[g(k;\hat{\underset{\sim}{\theta}}) - g(k;\underset{\sim}{\theta}^0)] = 2 &\sum_{t=1}^{\ell} (\hat{\theta}_t-\theta_t^0)\, g_t(k;\underset{\sim}{\theta}^0) \\ &+ \sum_{t=1}^{\ell}\sum_{s=1}^{\ell} (\hat{\theta}_t-\theta_t^0)(\hat{\theta}_s-\theta_s^0)\, g_{ts}(k;\underset{\sim}{\theta}^0) \\ &+ 2\beta_n \ell^3\, |\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}^0|_\ell^3\, G(y^k;\underset{\sim}{\theta}^0) . \end{aligned}$$

Summing from k=1,...,n on both sides we get

$$2[\log p(y^n|\hat{\underset{\sim}{\theta}}) - \log p(y^n|\underset{\sim}{\theta}^0)] = 2\sum_{t=1}^{\ell}\{(\hat{\theta}_t-\theta_t^0)\sum_{k=1}^{n} g_t(k;\underset{\sim}{\theta}^0)\}$$

$$+ \sum_{t=1}^{\ell}\sum_{s=1}^{\ell}\{(\hat{\theta}_t-\theta_t^0)\ (\hat{\theta}_s-\theta_s^0)\sum_{k=1}^{n} g_{ts}(k;\underset{\sim}{\theta}^0)\}$$

$$+ 2\beta_n \ell^3\ |\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}^0|_\ell^3 \sum_{k=1}^{n} G(y^k;\underset{\sim}{\theta}^0). \qquad (4.4.33)$$

Since

$$|\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}^0|_\ell^3 \sum_{k=1}^{n} G(y^k;\underset{\sim}{\theta}^0) = |\underset{\sim}{u}(n)|_\ell^3\ n^{-3/2}\sum_{k=1}^{n} G(y^k;\underset{\sim}{\theta}^0)$$

by (4.4.16), it follows, by (4.4.9) and (4.4.31) that

$$(\mathcal{L})\lim_{n\to\infty}\ [|\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}^0|_\ell^3 \sum_{k=1}^{n} G(y^k;\underset{\sim}{\theta}^0)] = 0 \qquad (4.4.34)$$

Now, (4.4.33) can be written as

$$-2\log \Lambda_n = 2\sum_{t=1}^{\ell} u_t(n)\ y_t(n)$$

$$+ \sum_{t=1}^{\ell}\sum_{s=1}^{\ell} u_t(n)\ u_s(n)\,[n^{-1}\sum_{k=1}^{n} g_{ts}(k;\underset{\sim}{\theta}^0)]$$

$$+ 2\beta_n \ell^3\ |\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}^0|_\ell^3 \sum_{k=1}^{n} G(y^k;\underset{\sim}{\theta}^0) \qquad (4.4.35)$$

where $\log \Lambda_n = \log p(y^n|\underset{\sim}{\theta}^0) - \log p(y^n|\hat{\underset{\sim}{\theta}})$. Therefore, by (4.4.7) - (4.4.9) and by (4.4.34) it follows that

$$(\mathcal{L})\lim_{n\to\infty} -2\log \Lambda_n = (\mathcal{L})\lim_{n\to\infty} < \Sigma_0\ \underset{\sim}{u}(n)\,,\ \underset{\sim}{u}(n) >.$$

Hence the required result follows from (4.4.31).

Now, in the following theorem, the conditions are obtained for strong consistency and asymptotic normality of an MLE when the n-dimensional joint density function depends on a single unknown parameter θ. When $\ell = 1$, the regularity conditions (E1) - (E7) reduce to the following form.

(C1) The parameter space Θ is an open interval in the real line.

(C2) For all k, $\frac{\partial^i}{\partial\theta^i} g(k,\theta)$, $i = 1,2$ exist for almost all y^k and for all $\theta \in \Theta$ and are continuous in θ.

(C3) For all k, for almost all y^{k-1} and for all $\theta \in \Theta$

$$E_\theta[\frac{\partial}{\partial\theta} g(k,\theta) | y^{k-1}] = 0, \tag{4.4.36}$$

and

$$-E_\theta[\frac{\partial^2}{\partial\theta^2} g(k,\theta) | y^{k-1}] = E_\theta[\{\frac{\partial}{\partial\theta} g(k,\theta)\}^2 | y^{k-1}] \tag{4.4.37}$$

$$= \sigma_k^2(\theta) \tag{4.4.38}$$

where $\sigma_k^2(\theta)$ is independent of $y_1, \ldots, y_{k-1}$.

(C4) $$\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} \sigma_k^2(\theta) = i(\theta) \tag{4.4.39}$$

where $i(\theta)$ is finite for all $\theta \in \Theta$. We shall further assume that $i(\theta)$ is continuous and non-zero for all $\theta \in \Theta$.

(C5) For all $\theta \in \Theta$ and for all k, there exists a neighbourhood $V(\theta)$, such that, for every $\theta' \in V(\theta)$

$$\left|\frac{\partial^2}{\partial\theta^2} g(k,\theta') - \frac{\partial^2}{\partial\theta^2} g(k,\theta)\right| \leq |\theta'-\theta| \, G(y^k,\theta) \tag{4.4.40}$$

where $G(y^k; \theta) \geq 0$,

and

$$(a.s)\lim_{n\to\infty} n^{-1}\sum_{k=1}^{n} E_{\theta}[G(y^{k},\theta)\,|\,y^{k-1}] = M(\theta) < \infty.$$

(C6) For all $\theta \varepsilon \Theta$

$$\text{(i)} \quad \sum_{k=1}^{\infty} k^{-2}\, \mathrm{Var}_{\theta}[\frac{\partial}{\partial\theta}\, g(k;\theta)] < \infty, \tag{4.4.41}$$

$$\text{(ii)} \quad \sum_{k=1}^{\infty} k^{-2}\, \mathrm{Var}_{\theta}[\frac{\partial^{2}}{\partial\theta^{2}}\, g(k,\theta)] < \infty, \tag{4.4.42}$$

and

$$\text{(iii)} \quad \sum_{k=1}^{\infty} k^{-2}\, \mathrm{Var}_{\theta}[G(y^{k},\theta)] < \infty. \tag{4.4.43}$$

(C7) For all $\theta \varepsilon \Theta$

$$\lim_{n\to\infty} n^{-(1+\delta/2)} \sum_{k=1}^{n} E_{\theta}\left|\frac{\partial}{\partial\theta} g(k,\theta)\right|^{2+\delta} = 0, \tag{4.4.44}$$

for some $\delta > 0$, δ independent of θ.

Now the following two theorems follow from Theorems 4.4.1 and 4.4.2.

Theorem 4.4.4: If assumptions (C1) - (C6) are satisfied, then the likelihood equation has a root which is strongly consistent.

Proof: Proof of this theorem follows from Theorem 4.4.1.

Theorem 4.4.5: Under the assumptions (C1) - (C7), a consistent MLE is asymptotically normal.

Proof: Proof follows from Theorem 4.4.2. In fact, if θ_o is the true parameter value of θ and $\hat{\theta}_n$ is an MLE, then it follows that

$$(\mathcal{L})\lim_{n\to\infty} n^{1/2}(\hat{\theta}_n-\theta_o) = N(0, i_o^{-1})$$

where $i_o = i(\theta_o)$.

In this case one can prove the following result:

Theorem 4.4.6: Under the assumptions (C1) - (C6), an MLE is asymptotically efficient of first-order.

Proof: In view of assumption (C4), the theorem follows as in Theorem 2.4.3.

4.5 The Bernstein-vonMises Theorem

Suppose that in addition to (C1) - (C7), the following conditions are also satisfied. Hereafter we write $h_k(\theta) \equiv g(k;\theta)$.

(C8) Λ is a prior probability measure on $(\Theta, \mathcal{F})$, where $\mathcal{F}$ is the σ-field of Borel subsets of Θ. Λ has a density λ with respect to the Lebesgue measure. The prior density λ is continuous and positive in an open neighbourhood of the true parameter θ_o.

(C9) Let, for each $\theta \in \Theta$ and any $\varepsilon > 0$,

$$R_k(\theta,\varepsilon) \equiv [\sup\{h_k(\theta') - h_k(\theta)\} : |\theta'-\theta| \geq \varepsilon, \theta' \in \Theta].$$

Then

$$(\text{a.s})\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} E\,[R_k(\theta,\varepsilon)|y^{k-1}]$$ is a finite negative quantity and

$$\sum_{k=1}^{\infty} k^{-2}\, \mathrm{Var}_\theta[R_k(\theta,\varepsilon)] < \infty.$$

(C10) Let θ_o denote the true parameter and $P_o = P_{\theta_o}$. Let H be a non-negative measurable function satisfying the following condition. There exists a number ε, $0 < \varepsilon < i_o$ such that

$$B(\theta_o) = (i_o/(2\pi))^{1/2} \int_{-\infty}^{\infty} H(t) \exp\{-(i_o-\varepsilon)t^2/2\} \, dt$$

is finite, and for every $h > 0$ and every $\delta > 0$

$$(a.s) \lim_{n\to\infty} [\exp(-\delta n) \int_{|t|>h} H(tn^{1/2}) \lambda(\hat{\theta}_n + t) dt] = 0$$

where $\hat{\theta}_n$ denotes an MLE.

We denote by $p_n(\theta|y^n) = p_n(\theta|y_1, \ldots, y_n)$, the posterior density based on the observations $y_1, \ldots, y_n$ corresponding to the prior probability density λ. Also, let

$$p_n^*(t|y^n) = n^{-1/2} \, p_n(\theta|y^n) \tag{4.5.1}$$

so that $p_n^*(t|y^n)$ is the posterior density of $n^{1/2}(\theta-\hat{\theta}_n)$. If $f(\theta)$ is a function differentiable twice with respect to θ, we denote the first and second derivatives of $f(\theta)$ by $f'(\theta)$ and $f''(\theta)$ respectively, in this section.

In Section 4.4, it was shown that a MLE $\hat{\theta}_n$ has the following properties, undder the assumptions (C1) - (C7)

(i) $$(a.s) \lim_{n\to\infty} \hat{\theta}_n = \theta_o, \tag{4.5.2}$$

(ii) $$\frac{\partial}{\partial\theta} \log p(y^n|\hat{\theta}_n) = 0 \quad \text{a.s for } n \geq N, \tag{4.5.3}$$

and

(iii) $$(\mathcal{L}) \lim_{n\to\infty} n^{1/2}(\hat{\theta}_n - \theta_o) = N(0, i_o^{-1}). \tag{4.5.4}$$

We shall now prove the following theorem which may be regarded as a generalisation of the Bernstein-vonMises theorem for dependent random variables. The proofs of the results of this section are similar to those of Borwanker, Kallianpur and Prakasa Rao [11]. We define

$$\nu_n(t) = \exp[\sum_{k=1}^{n} \{h_k(\hat{\theta}_n + tn^{-1/2}) - h_k(\hat{\theta}_n)\}] \tag{4.5.5}$$

and

$$C_n = \int_{-\infty}^{\infty} \nu_n(t)\ \lambda(\hat{\theta}_n + tn^{-1/2})\,dt. \tag{4.5.6}$$

Then from (4.5.1), it follows that

$$p_n^*(t|y^n) = C_n^{-1}\ \nu_n(t)\ \lambda(\hat{\theta}_n + tn^{-1/2}). \tag{4.5.7}$$

Theorem 4.5.1: Under assumptions (C1) - (C10) of Section 4.3,

$$(a.s) \lim_{n\to\infty} [\int_{-\infty}^{\infty} H(t)\,|p_n^*(t|y^n) - (i_o/2\pi)^{1/2}\exp(-i_o t^2/2)|\,dt] = 0.$$

The proof of this result is based on the following lemmas.

Lemma 4.5.1: Let assumptions (C1) - (C7) and (C8) be satisfied. Then the following conclusions hold:

(i) for every $\varepsilon(0 < \varepsilon < i_o)$, there exists a $\delta_o > 0$ and an integer N such that

$$\nu_n(t) \le \exp\{-(i_o - \varepsilon)t^2/2\} \tag{4.5.8}$$

for $|t| \le \delta_o n^{1/2}$ and $n \ge N$;

(ii) for every $\delta > 0$, there exists a positive ε and an integer N such that

$$\sup_{|t|>\delta n^{1/2}} \nu_n(t) \leq \exp(-n\varepsilon/4) \qquad (4.5.9)$$

for $n \geq N$; and

(iii) for every fixed t,

$$(a.s)\lim_{n\to\infty} \nu_n(t) = \exp(-i_o t^2/2). \qquad (4.5.10)$$

Proof: We have,

$$\log \nu_n(t) = \sum_{k=1}^{n} \{h_k(\hat{\theta}_n + tn^{-1/2}) - h_k(\hat{\theta}_n)\}$$

$$= tn^{-1/2} \sum_{k=1}^{n} h_k'(\hat{\theta}_n) + t^2/(2n) \sum_{k=1}^{n} h_k''(\theta_n') \qquad \ldots(4.5.11)$$

where $|\theta_n' - \hat{\theta}_n| \leq tn^{-1/2}$. The first term on the right hand side of (4.5.11) equals zero a.s for $n \geq N_1$ (say), by (4.5.3). For the second term we have

$$t^2/(2n) \sum_{k=1}^{n} h_k''(\theta_n') = t^2/(2n) \sum_{k=1}^{n} h_k''(\theta_o)$$

$$+ t^2/(2n) \sum_{k=1}^{n} \{h_k''(\theta_n') - h_k''(\theta_o)\}. \qquad (4.5.12)$$

Since the first term on the right hand side of (4.5.12) converges almost surely to $-i_o t^2/2$ by (C4), (C6) and Lemma 4.2.1 it follows that for a positive ε, ($\varepsilon < i_o$)

$$t^2/(2n) \sum_{k=1}^{n} h_k''(\theta_o) < (-i_o + \varepsilon/2)t^2/2 \qquad (4.5.13)$$

for $n \geq N_2$, where we may take $N_2 \geq N_1$. As $\hat{\theta}_n$ is a MLE which is strongly consistent by Theorem 4.4.1, we can choose a positive δ_o such that $0 < \delta_o < \varepsilon/(8M_o)$, $|\hat{\theta}_n-\hat{\theta}_\delta| < \delta_o$ and $|\theta_n'-\hat{\theta}_n| \leq t\, n^{-1/2} \leq \delta_o$ for $n \geq N_3$ (say). Hence, if $n \geq N_3$, $|\theta_n'-\theta_o| \leq 2\delta_o$ and by (C5),

$$n^{-1}\Big|\sum_{k=1}^{n}\{h_k''(\theta_n') - h_k''(\theta_o)\}\Big| \leq n^{-1}\, 2\delta_o \sum_{k=1}^{n} G(y^k;\theta_o) \quad (4.5.14)$$

which tends almost surely to $2\delta_o M_o$ by (4.4.11). Combining (4.5.11) - (4.5.14) we get

$$\log \nu_n(t) < -(i_o-\varepsilon)\, t^2/2 \quad (4.5.15)$$

for $|t| \leq n^{1/2}\,\delta_o$ and $n > N = \max(N_1, N_2, N_3)$. This proves (4.5.8). To prove (4.5.9), we have

$$n^{-1}\log \nu_n(t) = n^{-1}\sum_{k=1}^{n}\{h_k(\hat{\theta}_n+tn^{-1/2})-h_k(\theta_o)\}$$

$$+\, n^{-1}\sum_{k=1}^{n}\{h_k(\theta_o) - h_k(\hat{\theta}_n)\}. \quad (4.5.16)$$

If $|\hat{\theta}_n-\theta_o| < \delta/2$ for $n \geq N_4$, then $|tn^{-1/2}| > \delta$ implies that $|\hat{\theta}_n+tn^{-1/2}-\theta_o| > \delta/2$. Hence for $n \geq N_4$,

$$n^{-1}\sum_{k=1}^{n}\{h_k(\hat{\theta}_n+tn^{-1/2}) - h_k(\theta_o)\}$$

$$\leq n^{-1}\sum_{k=1}^{n}[\sup\{h_k(\theta)-h_k(\theta_o)\} : |\theta-\theta_o| > \delta/2]$$

$$= n^{-1}\sum_{k=1}^{n} R_k(\theta_o,\delta/2) \quad (4.5.17)$$

by (4.3.16). By (C9) and Lemma 4.2.1, we get that $n^{-1}\sum_{k=1}^{n} R_k(\theta_o, \delta/2)$ converges almost surely to a finite negative quantity. Therefore, for $n \geq N_4$,

$$n^{-1} \sum_{k=1}^{n} \{h_k(\hat{\theta}_n + tn^{-1/2}) - h_k(\theta_o)\} < 0. \tag{4.5.18}$$

Moreover,

$$n^{-1} \sum_{k=1}^{n} \{h_k(\theta_o) - h_k(\hat{\theta}_n)\} = (\theta_o - \hat{\theta}_n) n^{-1} \sum_{k=1}^{n} h_k'(\hat{\theta}_n)$$

$$+ (\theta_o - \hat{\theta}_n)^2/(2n) \sum_{k=1}^{n} h_k''(\theta_n') \tag{4.5.19}$$

which, by arguments given in (4.5.11) - (4.5.14) and by the fact that $\hat{\theta}_n$ converges a.s to θ_o as $n \to \infty$, converges a.s to zero as $n \to \infty$. Now choose ε such that

$$0 < \varepsilon < -(\text{a.s}) \lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} E_{\theta_o} [R_k(\theta_o, \delta/2) | y^{k-1}]. \tag{4.5.20}$$

Combining (4.5.16) and (4.5.18) - (4.5.20) it follows that,

$$\log \nu_n(t) \leq -n\varepsilon/4 \tag{4.5.21}$$

for $|t| > n^{1/2}\delta$ and $n \geq N = \max(N_1, N_2, N_3, N_4)$, which proves (4.5.9). Next, for a fixed t and $\varepsilon > 0$, choose an ε_1 such that $0 < \varepsilon_1 t^2/2 < \varepsilon$. By (4.5.11) - (4.5.14), we get that

$$|\log \nu_n(t) + t^2 i_o/2| < t^2 \varepsilon_1/2 \tag{4.5.22}$$

for $n \geq \max(N, t^2/\delta_o^2)$. (4.5.10) is thus proved.

In view of the above lemma, the proofs of the following two lemmas are the same as those of Lemma 3.2 and Lemma 3.3 of Borwanker, Kallianpur and Prakasa Rao [11]. But we give the proofs here for completeness.

Lemma 4.5.2: Under the assumptions (C1) - (C10), there exists a positive δ_o such that

$$(a.s)\lim_{n\to\infty}\left[\int_{|t|\le\delta_o n^{1/2}} H(t)\left|\nu_n(t)\lambda(\hat\theta_n+tn^{-1/2})-\lambda(\theta_o)\exp(-i_o t^2/2)\right|dt\right] = 0 .$$

Proof: It is easy to see that

$$\int_{|t|\le\delta_o n^{1/2}} H(t)\left|\nu_n(t)\lambda(\hat\theta_n+tn^{-1/2})-\lambda(\theta_o)\exp(-i_o t^2/2)\right|dt$$

$$\le \int_{|t|\le\delta_o n^{1/2}} H(t)\lambda(\theta_o)\left|\nu_n(t)-\exp(-i_o t^2/2)\right|dt$$

$$+ \int_{|t|\le\delta_o n^{1/2}} H(t)\nu_n(t)\left|\lambda(\theta_o)-\lambda(\hat\theta_n+tn^{-1/2})\right|dt. \quad (4.5.23)$$

Choose an $\varepsilon > 0$ such that

$$\int_{-\infty}^{\infty} H(t)\ \exp[-(i_o-\varepsilon)t^2/2]dt$$

is finite. This is possible because of (C10). Then there exist a δ_1 and an N such that

$$\nu_n(t) \le \exp\ [-(i_o-\varepsilon)t^2/2]$$

for $|t| \le \delta_1 n^{1/2}$ and $n \ge N$, by Lemma 4.5.1. Hence, by (4.5.22), we have

$$(a.s)\lim_{n\to\infty}\left[\int_{|t|\le\delta_1 n^{1/2}} H(t)\lambda(\theta_o)\left|\nu_n(t)-\exp(-i_o t^2/2)\right|dt\right] = 0$$

by the dominated convergence theorem. Further,

$$\int_{|t|\leq\delta_1 n^{1/2}} H(t)\nu_n(t)|\lambda(\theta_o)-\lambda(\hat{\theta}_n+tn^{-1/2})|\,dt$$

$$\leq \sup_{|\theta-\theta_o|\leq\delta_1} |\lambda(\theta)-\lambda(\theta_o)|\,[\int_{|t|\leq\delta_1 n^{1/2}} H(t)\exp[-(i_o-\varepsilon)t^2/2]dt].$$

For a given η, choose $\delta_o \leq \delta_1$ such that

$$\sup_{|\theta-\theta_o|<\delta_o} |\lambda(\theta)-\lambda(\theta_o)| \int_{|t|\leq\delta_o n^{1/2}} H(t)\exp[-(i_o-\varepsilon)t^2/2]dt < \eta . \qquad (4.5.24)$$

Combining (4.5.23) and (4.5.24), we get the required lemma.

Lemma 4.5.3: Under assumptions (C1) - (C10), for every $\delta > 0$,

$$(a.s)\lim_{n\to\infty} [\int_{|t|>\delta n^{1/2}} H(t)|\nu_n(t)\lambda(\hat{\theta}_n+tn^{-1/2})-\lambda(\theta_o)\exp(-i_o t^2/2)|dt] = 0.$$

Proof: It is clear that the term in brackets on the left side is

$$\leq \int_{|t|>\delta n^{1/2}} H(t)\ \nu_n(t)\ \lambda(\hat{\theta}_n+tn^{-1/2})\ dt$$

$$+ \int_{|t|>\delta n^{1/2}} H(t)\ \lambda(\theta_o)\ \exp(-i_o t^2/2)\ dt$$

$$\leq \exp(-n\varepsilon/4) \int_{|t|>\delta n^{1/2}} H(t)\ \lambda(\hat{\theta}_n+tn^{-1/2})\ dt$$

$$+ \lambda(\theta_o) \int_{|t|>\delta n^{1/2}} H(t)\ \exp(-i_o t^2/2)\ dt \qquad (4.5.25)$$

by (4.5.9). By assumption (C10), the right side of (4.5.25) tends to zero a.s as $n \to \infty$. Hence the lemma follows.

Now, we prove Theorem 4.5.1. From Lemmas 4.5.2 and 4.5.3, we obtain that

$$(a.s)\lim_{n\to\infty} \int_{-\infty}^{\infty} H(t)\,|\nu_n(t)\lambda(\hat{\theta}_n+tn^{-1/2})-\lambda(\theta_o)\exp(-i_o t^2/2)|\,dt = 0. \tag{4.5.26}$$

Putting $H(t) \equiv 1$, which satisfies the assumptions on the function H trivially, we get that

$$\begin{aligned} C_n &= \int_{-\infty}^{\infty}\nu_n(t)\;\lambda(\hat{\theta}_n+tn^{-1/2})\;dt \to \lambda(\theta_o)\int_{-\infty}^{\infty}\exp(-i_o t^2/2)\;dt \\ &= \lambda(\theta_o)\;(2\pi/i_o)^{1/2}. \end{aligned} \tag{4.5.27}$$

Hence, by (4.5.6),

$$\begin{aligned} &\int_{-\infty}^{\infty} H(t)\,|p_n^*(t|y^n)-(i_o/(2\pi))^{1/2}\;\exp(-i_o t^2/2)|\,dt \\ &\le \int_{-\infty}^{\infty} H(t)\,|C_n^{-1}\;\lambda(\hat{\theta}_n+tn^{-1/2})\nu_n(t)-C_n^{-1}\lambda(\theta_o)\exp(-i_o t^2/2)|\,dt \\ &\quad + \int_{-\infty}^{\infty} H(t)\,|C_n^{-1}\;\lambda(\theta_o)-(i_o/(2\pi))^{1/2}|\exp(-i_o t^2/2)\,dt. \end{aligned}$$

Hence Theorem 4.5.1 follows by (4.5.26) and (4.5.27).

As a corollary to Theorem 4.5.1, we give below a result which includes, besides the more traditional form of the theorem of Bernstein-vonMises, other interesting variations.

Theorem 4.5.2: Let assumptions (C1) - (C7) be satisfied. Let the prior density λ satisfy assumption (C8) and

$$\int_{-\infty}^{\infty} |\theta|^m\lambda(\theta)\;d\theta < \infty \tag{4.5.28}$$

for some non-negative integer m. Then

$$(a.s)\lim_{n\to\infty} \int_{-\infty}^{\infty} |t|^m \, |p_n^*(t|y^n) - (i_o/(2\pi))^{1/2} \exp(-i_o t^2/2)| dt = 0. \tag{4.5.29}$$

Proof: For m $\geq$ 1, set $H(\theta) = |\theta|^m$ in Theorem 4.5.1. From C_r-inequality, we have for positive numbers h and δ

$$e^{-\delta n} \int_{|t|>h} H(tn^{1/2}) \, \lambda(\hat{\theta}_n+t) \, dt$$

$$= n^{m/2} \, e^{-\delta n} \int_{|t-\hat{\theta}_n|>h} \lambda(t) |t-\hat{\theta}_n|^m \, dt$$

$$\leq 2^{m-1} \, n^{m/2} \, e^{-\delta n} \left[\int_{|t-\hat{\theta}_n|>h} \lambda(t)|t|^m \, dt + |\hat{\theta}_n|^m \int_{|t-\hat{\theta}_n|>h} \lambda(t) dt \right]$$

$$\leq 2^{m-1} \, n^{m/2} \, e^{-\delta n} \left[\int_{-\infty}^{\infty} |t|^m \lambda(t) dt + |\hat{\theta}_n|^m \right] \tag{4.5.30}$$

which tends to zero a.s by Theorem 4.4.1 and assumption (4.5.28). It easily follows that $B(\theta_o)$ in (C10) is finite for $H(t) = |t|^m$. Hence (C10) is satisfied. Thus the conclusion of Theorem 4.5.2 follows from Theorem 4.5.1.

For m = 0 the assertion of Theorem 4.5.2 is the classical form of the Bernstein-vonMises theorem. Theorem 4.5.2 is an extension to arbitrary stochastic process of the corresponding result for the case of Markov processes due to Borwanker, Kallianpur and Prakasa Rao [11].

As an application of Theorem 4.5.1 to the theory of asymptotic Bayesian inference for stochastic process, we shall derive results similar to those in Borwanker, Kallianpur Kallianpur and Prakasa Rao [11].

Following LeCam [25], we define a regular Bayes estimate $T_n = T_n(y^n)$ as an estimate which minimizes

$$B_n(\beta) = \int \tilde{\ell}(\theta,\beta)\ p_n(\theta|y^n)\ d\theta$$

for all $(y_1,\ldots,y_n)$ and all n, where $\tilde{\ell}(\theta,\beta)$ is a loss function defined on $\Theta\times\Theta$. We assume that a measurable, regular Bayes estimate T_n of θ exists.

Theorem 4.5.3: Let $\{Y_n, n \geq 1\}$ be a process satisfying assumptions (C1) - (C10). Let T_n be a measurable regular Bayes estimate of θ with respect to a loss function $\tilde{\ell}(\theta,\beta)$ which satisfies the following conditions.

(i) $\tilde{\ell}(\theta,\beta) = \ell(\theta-\beta) \geq 0$.

(ii) $\ell(t_1) \geq \ell(t_2)$ if $t_1 > t_2 \geq 0$ or if $t_1 < t_2 \leq 0$.

There exist constants $\{a_n\}$ and functions $K(t)$ and $G(t)$ such that

(iii) $a_n \geq 0$,

(iv) $G(t)$ satisfies assumption (C10) and $a_n\ \ell(tn^{-1/2}) \leq G(t)$ for all n,

(v) $a_n\ \ell(tn^{-1/2}) \to K(t)$ uniformly on compact sets, as $n\to\infty$.

(vi) $\int K(t+m)\exp(-i_o t^2/2)\ dt$ has a strict minimum at $m = 0$.

Then

(i) $$(a.s)\lim_{n\to\infty} (T_n - \theta_o) = 0, \qquad (4.5.31)$$

(ii) $$(\mathcal{L})\lim_{n\to\infty} n^{1/2}(\theta_o - T_n) = N(0,\ i_o^{-1}) \ , \qquad (4.5.32)$$

and

(iii) $$(a.s)\lim_{n\to\infty} a_n\ B_n(T_n) = (i_o/(2\pi))^{1/2}\int K(t)\ \exp(-i_o t^2/2)dt.$$

Proof: In view of Theorem 4.5.1, the proof of this theorem is the same as that of Theorem 4.1 of Borwanker, Kallianpur and Prakasa Rao [11].

Remark 4.4.3: Chao [13] showed that the maximum likelihood estimator $\hat{\underset{\sim}{\theta}}_n$ and the Bayes estimator $\bar{\underset{\sim}{\theta}}_n$ are asymptotically equivalent under some assumptions on the likelihood function and the loss function, when the observations are independent and identically distributed. Similar results can be obtained for arbitrary stochastic processes satisfying conditions of Section 4.3.

4.6 Examples:

In this section, we present two examples which satisfy the conditions (C1) - (C7) of Section 4.3 and thus establish strong consistency, asymptotic normality and first-order efficiency of MLEs of unknown parameters. The first example is a first-order auto-regressive scheme, which forms a non-stationary Markov process and the second example is a non-stationary, equi-correlated stochastic process.

Example 4.6.1: Let the sequence of random variables $\{X_n, n \geq 1\}$ satisfy the relation

$$X_n = \alpha X_{n-1} + u_n \tag{4.6.1}$$

where $|\alpha| < 1$ is known and the sequence of random variables $\{u_n, n \geq 1\}$ be independent, normally distributed with unknown mean μ and variance unity. Further, let $u_n \equiv 0$ for $n \leq 0$. From (4.6.1) we get

$$X_n = \sum_{i=1}^{n} \alpha^{i-1} u_{n-i+1} \tag{4.6.2}$$

so that

$$E X_n = \mu(1-\alpha^n)(1-\alpha)^{-1} \tag{4.6.3}$$

and for $n \geq m$

$$\mathrm{Cov}(X_n, X_m) = \alpha^{n-m}(1-\alpha^{2m})(1-\alpha^2)^{-1}. \tag{4.6.4}$$

Since the $n \times n$ variance-covariance matrix Γ_n of $(X_1, \ldots, X_n)$ is given by (4.6.4), we have

$$\Gamma_n^{-1} = \begin{bmatrix} 1+\alpha^2 & -\alpha & 0 & \cdots & 0 & 0 \\ -\alpha & 1+\alpha^2 & -\alpha & \cdots & 0 & 0 \\ 0 & -\alpha & 1+\alpha^2 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -\alpha & 1 \end{bmatrix} \tag{4.6.5}$$

Let Γ_{n-1} denote the variance-covariance matrix of the $1 \times (n-1)$ random vector $\underset{\sim}{X}^{(n-1)} \equiv \{X_1, \ldots, X_{n-1}\}$ and $\Gamma_{1,n-1}$ be the $1\times(n-1)$ vector $\{\alpha^{n-1}(1-\alpha^2),\ \alpha^{n-2}(1-\alpha^4),\ldots,\alpha(1-\alpha^{2n-2})\}$. By Theorem 2.5.1 of Anderson [2], the conditional density function of X_n given $\underset{\sim}{X}^{(n-1)} = \underset{\sim}{x}^{(n-1)}$ is

$$p_n(\mu) = (2\pi)^{-1/2}\ D_n^{-1}\ \exp[-(2d_n^2)^{-1}\ (x_n-C_n)^2] \qquad (4.6.6)$$

where

$$C_n = E(X_n) + (1-\alpha^2)^{-1}\ \Gamma_{1,n-1}\ \Gamma_{n-1}^{-1}(\underset{\sim}{x}^{(n-1)} - \underset{\sim}{\varepsilon}^{(n-1)})^T, \qquad \ldots(4.6.7)$$

$$\underset{\sim}{\varepsilon}^{(n-1)} = \{E(X_1),\ E(X_2),\ldots,\ E(X_{n-1})\}, \qquad (4.6.8)$$

and

$$D_n^2 = \mathrm{Var}(X_n) - (1-\alpha^2)^{-2}\ \Gamma_{1,n-1}\ \Gamma_{n-1}^{-1}\ \Gamma_{1,n-1}^T. \qquad (4.6.9)$$

It is clear that,

$$(1-\alpha^2)^{-2}\ \Gamma_{1,n-1}\ \Gamma_{n-1}^{-1}\ \Gamma_{1,n-1}^T = \alpha^2(1-\alpha^{2n-2})(1-\alpha^2)^{-1} \qquad \ldots(4.6.10)$$

and

$$C_n = r_n + (1-\alpha^2)^{-1}\ \Gamma_{1,n-1}\ \Gamma_{n-1}^{-1}\ (\underset{\sim}{x}^{(n-1)})^T \qquad (4.6.11)$$

where

$$\begin{aligned} r_n &= \mu(1-\alpha)^{-1}\ (1-\alpha^n) - (1-\alpha^2)^{-1}\ \Gamma_{1,n-1}\ \Gamma_{n-1}^{-1}(\underset{\sim}{\varepsilon}^{(n-1)})^T \\ &= \mu(1-\alpha)^{-1}\ (1-\alpha^n) - \alpha E(X_{n-1}) \\ &= \mu(1-\alpha)^{-1}\ (1-\alpha^n) - \alpha\mu(1-\alpha)^{-1}\ (1-\alpha^{n-1}) = \mu\ . \end{aligned} \qquad (4.6.12)$$

Therefore, by (4.6.8) - (4.6.12), it follows that

$$C_n = \mu + \alpha\ x_{n-1}\ , \qquad (4.6.13)$$

and

$$D_n^2 = 1\ . \qquad (4.6.14)$$

Therefore,

$$p_n(\mu) = (2\pi)^{-1/2} \exp[-(x_n-\mu-\alpha x_{n-1})^2/2] \qquad (4.6.15)$$

so that

$$\frac{\partial}{\partial\mu} \log p_n(\mu) = x_n - \mu - \alpha x_{n-1} \qquad (4.6.16)$$

and

$$\frac{\partial^2}{\partial\mu^2} \log p_n(\mu) = -1 \ . \qquad (4.6.17)$$

Assumptions (C1) and (C2) follow trivially and to verify assumption (C3), we observe that

$$E_\mu[\frac{\partial}{\partial\mu} \log p_n(\mu) \mid x_1,\ldots,x_{n-1}]$$

$$= E_\mu[X_n-(\mu+x_{n-1})|x_1,\ldots,x_{n-1}] = 0 \qquad (4.6.18)$$

and

$$E_\mu[\{\frac{\partial}{\partial\mu}\log p_n(\mu)\}^2 \, |x_1,\ldots,x_{n-1}]$$

$$= E_\mu[\{X_n-(\mu+x_{n-1})\}^2 \, |x_1,\ldots, x_{n-1}]$$

$$= D_n^2 = 1 \qquad (4.6.19)$$

by (4.6.13) and (4.6.14). Thus (C3) and hence (C4) follows from (4.6.19) with $i(\mu) = 1$ for all μ. Since $\frac{\partial^2}{\partial\mu^2} \log p_n(\mu)$ is independent of $x_1,\ldots,x_{n-1}$, (C5) will be satisfied with $G(x_1,\ldots,x_n;\mu) \equiv 0$. By (4.6.18), (4.6.19) and by assumptions (C4) and (C5), the validity of (C6) follows. The conditional distribution of $\frac{\partial}{\partial\mu} \log p_n(\mu)$ given $\underset{\sim}{X}^{(n-1)} = \underset{\sim}{x}^{(n-1)}$ is normal with mean $\mu+x_{n-1}$ and variance one by (4.6.15). Therefore

$$E_\mu[|\frac{\partial}{\partial\mu} \log p_n(\mu)|^3 \, |\underset{\sim}{X}^{(n-1)} = \underset{\sim}{x}^{(n-1)}] = 2\sqrt{2/\pi}. \qquad (4.6.20)$$

It follows from (4.6.20) that (C7) is satisfied. Thus all the assumptions of Theorems 4.4.4, 4.4.5 and 4.4.6 are satisfied and hence the MLE of μ is strongly consistent, asymptotically normal and first-order efficient.

Example 4.6.2: Let $\{X_n, n \geq 1\}$ be a sequence of random variables such that (i) $E(X_n) = \mu$ or 2μ according as n is odd or even; (ii) $\text{Cov}(X_n, X_m) = C$ for all n and m $(n \neq m)$, and (iii) $\text{Var}(X_n) = 2C$ for all n. Further we suppose that C is known and μ is unknown. The problem is to obtain the asymptotic properties of a MLE of μ. Suppose that, for all n, the n-dimensional joint density function is given by $p(x_1,\ldots,x_n;\mu)$

$$= (2\pi)^{-n/2} \{\det \Gamma_n\}^{1/2} \exp[-\{(\underset{\sim}{x}^{(n)} - \underset{\sim}{\varepsilon}^{(n)})\Gamma_n^{-1}(\underset{\sim}{x}^{(n)} - \underset{\sim}{\varepsilon}^{(n)})^T\}/2] \qquad \ldots (4.6.21)$$

where $\underset{\sim}{x}^{(n)}$ and $\underset{\sim}{\varepsilon}^{(n)}$ are $1\times n$ vectors $\{x_1,\ldots,x_n\}$ and $\{E(X_1),\ldots, E(X_n)\}$ respectively, and Γ_n^{-1} is the inverse of the variance-covariance matrix Γ_n of the random vector $\underset{\sim}{X}^{(n)}$. Here

$$\Gamma_n^{-1} = \{C(n+1)\}^{-1} \begin{bmatrix} n & -1 & -1 & \ldots & -1 \\ -1 & n & -1 & \ldots & -1 \\ .. & .. & .. & \ldots & .. \\ .. & .. & .. & \ldots & .. \\ -1 & -1 & -1 & \ldots & n \end{bmatrix} \qquad (4.6.22)$$

As in Example 4.6.1, by Theorem 2.5.1 of Anderson [2], we get that the conditional density of X_n given

$\underset{\sim}{X}^{(n-1)} = \underset{\sim}{x}^{(n-1)}$ to be

$$p_n(\mu) = (2\pi)^{-n/2} D_n^{-1} \exp[-\{2D_n^2\}^{-1} (x_n - c_n)^2] \qquad (4.6.23)$$

where

$$c_n = E(X_n) + n^{-1} \sum_{i=1}^{n-1} \{x_i - E(X_i)\} \qquad (4.6.24)$$

and

$$D_n^2 = \mathrm{Var}(X_n) - n^{-1} C(n-1). \qquad (4.6.25)$$

Hence, by (i), (ii) and (iii), we get that

$$c_n = \begin{cases} (2n)^{-1} (8-n)\mu + n^{-1} \sum_{i=1}^{n-1} x_i & \text{if } n \text{ is even} \\ (2n)^{-1} (5-3n)\mu + n^{-1} \sum_{i=1}^{n-1} x_i & \text{if } n \text{ is odd} \end{cases} \qquad (4.6.26)$$

and

$$D_n^2 = C\, n^{-1} (n+1). \qquad (4.6.27)$$

Therefore,

$$\log p_n(\mu) = -\frac{1}{2} \log 2\pi - \frac{n}{2C(n+1)} (x_n - r_n\mu - n^{-1} \sum_{i=1}^{n-1} x_i)^2$$

$$\ldots (4.6.28)$$

where $r_n = (2n)^{-1}(8-n)$ or $(2n)^{-1}(5-3n)$ according as n is even or odd. Since

$$\frac{\partial}{\partial\mu} \log p_n(\mu) = \frac{nr_n}{C(n+1)} (x_n - r_n\mu - n^{-1} \sum_{i=1}^{n-1} x_i) \qquad (4.6.29)$$

and

$$\frac{\partial^2}{\partial\mu^2} \log p_n(\mu) = -\frac{nr_n^2}{C(n+1)}. \qquad (4.6.30)$$

Assumptions (C1) - (C4) and (C6) are easy to verify and follow as in Example 4.6.1. Since $\frac{\partial^2}{\partial\mu^2}\log p_n(\mu)$ is independent of $x_1,\ldots,x_{n-1}$, (C5) will be satisfied with $G(x_1,\ldots,x_n;\mu) \equiv 0$. As in Example 4.6.1, (C7) follows. From (4.6.30) we get

$$\sigma_k^2 = E[\frac{\partial}{\partial\mu}\log p_n(\mu)]^2 = \frac{r_k^2\, k}{C(k+1)} \quad .$$

Hence σ_k^2 are uniformly bounded by 4/C. Therefore, $i(\theta) = \lim_{n\to\infty} n^{-1}\sum_{k=1}^{n}\sigma_k^2$ is finite if it exists. In fact it can be shown in the following way that $i(\theta) = 1/4C$.

We have

$$- I_n(\mu) = E_\mu[\frac{\partial^2}{\partial\mu^2}\log p(x_1,\ldots,x_n;\mu)$$

$$= \sum_{k=1}^{n} E_\mu \frac{\partial^2}{\partial\mu^2}\log p_k(\mu)$$

$$= \sum_{k=1}^{n} E_\mu[E_\mu \frac{\partial^2}{\partial\mu^2}\log p_k(\mu)\,|\,x_1,\ldots,x_{k-1}]$$

$$= \sum_{k=1}^{n} E_\mu[-\sigma_k^2] .$$

Since σ_k^2 is independent of $x_1,\ldots,x_{k-1}$, we get

$$I_n(\mu) = \sum_{k=1}^{n}\sigma_k^2 \quad .$$

Therefore,

$$\lim_{n\to\infty} n^{-1}\sum_{k=1}^{n}\sigma_k^2 = \lim_{n\to\infty} n^{-1} I_n(\mu) \quad . \qquad (4.6.31)$$

We have $\log p(x_1,\ldots,x_n;\mu)$

$$= -\frac{n}{2}\log 2\pi - \frac{1}{2}\log(\det\Gamma_n) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_n^{ij}(x_i-a_i\mu)(x_j-a_j\mu) \qquad \ldots(4.6.32)$$

where γ_n^{ij} are the elements of Γ_n^{-1} and $\dot{a}_n$ is equal to 1 or 2 according as n is odd or even. From (4.6.32) we get

$$\frac{\partial^2}{\partial\mu^2} \log p(x_1,\ldots,x_n;\mu) = - \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_n^{ij} a_i a_j .$$

After further simplification, it can be shown that

$$I_n(\mu) = \begin{cases} n(n+5)\{4C(n+1)\}^{-1} & \text{if n is even} \\ (n^2+10n-7)\{4C(n+1)\}^{-1} & \text{if n is odd.} \end{cases}$$

Hence it follows that

$$i(\mu) = \lim_{n\to\infty} n^{-1} I_n(\mu) = 1/4C.$$

Thus all the assumptions of Theorems 4.4.4, 4.4.5 and 4.4.6 are satisfied and hence it follows that the maximum likelihood likelihood estimator $\hat{\mu}_n$ of μ is strongly consistent, asymptotically normal and **first-order** efficient.

4.7 Conclusions:

In this chapter, we obtained strong consistency, asymptotic normality and first-order efficiency of MLE when the observations are from arbitrary stochastic process. m-dependence and ϕ-mixing treated in the previous two chapters are special cases of stochastic processes, but the set of conditions of Section 4.3 of this chapter, will not imply the conditions assumed for them. However, the examples presented in the previous two chapters will satisfy the assumptions of this chapter.

CHAPTER V

COMMENTS AND SUGGESTIONS FOR FURTHER WORK

In the previous chapters, the consistency either strong or weak, asymptotic normality and asymptotic efficiency of first-order are proved when the observations are from dependent random variables. When the observations are independent, all the three sets of conditions presented in Sections 2.3, 3.5 and 4.3 reduce to Cramér's [14] conditions, of which some may be superfluous. Thus we note that the results obtained in these pages are natural extensions to dependent random variables, of the existing results for independent random variables. Except for this fact, the results of Chapter II and Chapter III cannot be compared with any other results as there are none in this direction till now. However, the results about consistency in Chapter III may be compared with those of Bar-Shalom [4] for arbitrary stochastic processes. These results may be further improved as pointed out below.

The regularity conditions presented in Sections 2.3, 3.5 and 4.3 may be relaxed so as to cover a wider class of density functions.

Conditions may be established which ensure strong consistency and asymptotic normality of MLE in the m-dependent case and strong consistency in ϕ-mixing case.

Ibragimov [20] has proved the following theorem: For a sequence $\{X_j, j \geq 1\}$ of stationary ϕ-mixing random variables, if $\lim_{n\to\infty} \phi_n = 0$, $E(X_1) = 0$, $E|X_1|^{2+\delta} < \infty$ for some $0 < \delta < 1$ and $\lim_{n\to\infty} E(\sum_{i=1}^{n} X_i)^2 = \infty$, then there exists a constant C such that

$$E\left|\sum_{i=1}^{n} X_i\right|^{2+\delta} \leq C\{E(\sum_{i=1}^{n} X_i)^2\}^{1+\delta/2} .$$

The extension of this result to double sequences of stationary ϕ-mixing random variables is very much useful. If this result is obtained, the assumption on the rate of convergence to zero of the mixing coefficient can be removed. Moreover, the proofs of the results of Chapter III can be simplified.

Since theorems on asymptotic behaviour of estimators are useful for practical purposes, only when the rates of convergence are known, some work in this area has been done by Pfanzagl [33] in the independent case and Prakasa Rao [36] for Markov processes. Similar results may be obtained for the general cases discussed in the previous chapter.

Billingsley [8] and Sarma [45] studied some asymptotic properties of the test statistics where the random variables are Markov dependent. Similar results have to be obtained for the cases treated in the previous chapters.

REFERENCES

[1] AITCHISON, J., and SILVEY, S.D. (1958). Maximum likelihood estimation of parameters subject to restraints. Ann. Math. Statist. 29, 813-828.

[2] ANDERSON, T.W. (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York.

[3] BAHADUR, R.R. (1960). On the asymptotic efficiency of tests and estimators, Sankhya. 22, 229-252.

[4] BAR-SHALOM, Y. (1971). On the asymptotic properties of maximum likelihood estimate obtained from dependent observations. J. Roy. Statist.Soc.Ser.B. 33. 72-77.

[5] BHAT, B.R., and KULKARNI, S.R. (1972). Optimal asymptotic tests of composite hypotheses for stochastic processes. (Preprint).

[6] BICKEL, P. and YAHAV, J. (1969). Some contributions to the asymptotic theory of Bayes solutions. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete. II, 257-275.

[7] BILLINGSLEY, P. (1961). Levy-Lindeberg theorem for martingales. Proc. Amer.Math.Soc. 12, 788-792.

[8] BILLINGSLEY, P. (1961). Statistical inference for Markov processes. University Chicago Press.

[9] BILLINGSLEY, P. (1965). Ergodic theory and information. Wiley, New York.

[10] BILLINGSLEY, P. (1968). Convergence of probability measures. Wiley, New York.

[11] BORWANKER, J., KALLIANPUR, G., and PRAKASA RAO, B.L.S. (1971). The Bernstein-vonMises theorem for Markov processes. Ann. Math. Statist. 42, 1241-1253.

[12] CHAO, M.T. (1970). Strong consistency of maximum likelihood estimators when the observations are independent but not identically distributed. Dr.Y.W. Chen's 60-year Memorial Volume. Academia Sinica, Taipei.

[13] CHAO, M.T. (1970). The asymptotic behavior of Bayes estimators. *Ann. Math.Statist.* **41**, 601-608.

[14] CRAMER, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press, Princeton.

[15] CRAMER, H., and WOLD, H. (1936). Some theorems on distribution functions. *J. Lond.Math.Soc.* **11**, 290-295.

[16] DOOB, J.L. (1953). *Stochastic Processes.* Wiley, New York.

[17] GNEDENKO, B.V., and KOLMOGROV, A.N. (1954). *Limit distributions for sums of independent random variables.* Addison-Wesley, Cambridge, Mass.

[18] HOADLEY, B. (1971). Asymptotic properties of likelihood estimators for the independent but not identically distributed case. *Ann. Math. Statist.* **42**, 1977-1991.

[19] HUBER, P.J. (1967). The behaviour of maximum likelihood estimates under non-standard conditions. *Proc. 5th Berkeley Symp. Math.Stat.Prob.* **1**, 221-233.

[20] IBRAGIMOV, I.A. (1962). Some limit theorems for stationary processes. *Theor. Probab. Appl.* **7**, 349-381.

[21] IBRAGIMOV, I.A., and HAS'MINSKII, R.Z. (1970). Asymptotic behaviour of generalised Bayes estimates. *Soviet Math. Dokl.* **11**, 1181-1185.

[22] IBRAGIMOV, I.A., and HAS'MINSKII, R.Z. (1970). On the limiting behaviour of Bayes estimates and maximum likelihood estimates. *Soviet Mach.Dokl.* **12**, 831-834.

[23] IBRAGIMOV, I.A., and HAS'MINSKII, R.Z. (1972). Asymptotical behavioud of some statistical estimators in the smooth case I. Studying the likelihood ratio. *Teoriya Veroyatnostei i ee Primeneniya* (Russian). **17**, 469-486.

[24] LeCAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. of California Publications in Statistics.* **1**, 277-330.

[25] LeCAM, L. (1955). On the asymptotic theory of estimation and testing of hypotheses. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1**, 129-156.

[26] LeCAM, L. (1958). Les proprietes asymptotiques des solutions de Bayes. *Publ. Inst.Statist.Univ.Paris.* **7**, 17-35.

[27] LINDLEY, D.V. (1961). The use of prior probability distributions in statistical inference and decision. *Proc. 4th Berkeley Symp. Math.Statist.Prob.* **1**, 435-468.

[28] LOÉVE, M. (1963). *Probability Theory*. 3rd edition. Van Nostrand, Princeton.

[29] MORAN, P.A.P. (1971). The uniform consistency of maximum-likelihood estimators. *Proc. Camb.Phil.Soc.* **70**, 435-440.

[30] MORAN, P.A.P. (1971). Maximum likelihood estimation in non-standard conditions. *Proc. Camb. Phil.Soc.* **70**, 441-450.

[31] NOCTURNE, D. Asymptotic efficiency of the maximum likelihood estimators for the parameters of certain stochastic processes. *Ph.D. Thesis*, *Technical Report* no.105, Department of *Operations* Research, Cornell University.

[32] PARZEN, E. (1960). *Modern Probability theory and its applications*. Wiley, New York.

[33] PFANZAGL, J. (1971). The Berry-Eseen bound for minimum contrast estimates. *Metrika*, **17**, 81-91.

[34] PRAKASA RAO, B.L.S. (1968). Estimation of the location of the cusp of a continuous density. *Ann. Math.Statist.* **39**, 76-87.

[35] PRAKASA RAO,B.L.S. (1972). Maximum likelihood estimation for Markov processes. *Ann. Inst.Stat.Math.* (To appear).

[36] PRAKASA RAO, B.L.S. (1973). On the rate of convergence of estimators for Markov processes. *Z.Wahrscheinlichkeitstheorie und Verw.Gebiete.* (To appear).

[37] RAO, C.R. (1961). Asymptotic efficiency and limiting information. *Proc. Fourth Berkeley Symp. Math.Statist. Prob.* **1**, 531-546.

[38] RAO, M.M. (1961). Consistency and limit distributions of estimators of aparameters in explosive stochastic difference equations. *Ann. Math.Statist.* **32**, 195-218.

[39] RAO, M.M. (1966). Inference in stochastic processes,II. *Z.Wahrscheinlichkeitstheorie und Verw. Gebiete.* **5**, 317-335.

[40] ROSEN, B. (1967). On the central limit theorem for sums of dependent random variables. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete. 7, 48-82.

[41] ROSENBLATT, M. (1962). Random Processes. Oxford Univ. Press, New York.

[42] ROUSSAS, G.G. (1965). Extension to Markov processes of a result by A. Wald about the consistency of the maximum likelihood estimate. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete. 4, 69-73.

[43] ROUSSAS, G.G. (1968). Asymptotic normality of the maximum likelihood estimate in Markov Processes. Metrika. 14, 62-70.

[44] ROUSSAS, G.G.m(1972). Contiguity of probability measures. Some applications in Statistics. Cambridge University Press.

[45] SARMA, Y.R.K. (1968). Sud les tests et sur L'estimation de parametres pour certains processus stochastiques stationaires. Publ.Inst.Statist.Univ.Paris. 17, 1-124.

[46] SERFLING, R.J.(1968). Contributions to central limit theory for dependent variables. Ann. Math.Statist. 39 1158-1175.

[47] SHAMAN, P. (1969). On the inverse of the covariance matrix of a first-order moving average. Biometrika. 56, 595-600.

[48] SILVEY, S.D. (1961). A note on Maximum likelihood in the case of dependent random variables. J. Roy. Statist. Soc.Ser.B. 23, 444-452.

[49] TUCKER, H.G. (1967). A Graduate course in probability. Academic Press New York.

[50] WALD, A. (1948). Asymptotic properties of the maximum likelihood estimate of an unknown parameter of a discrete stochastic process. Ann. Math.Statist.19, 40-40-46.

[52] WOLFOWITZ, J. (1949). On Wald's proof of the consistency of the maximum likelihood estimate. Ann. Math. Statist. 20, 601-602.

[53] ZACKS, S. (1971). The theory of statistical inference. Wiley, New York.